“华为杯”第十五届中国研究生
数学建模竞赛

题 目　　对恐怖袭击事件记录数据的量化分析

摘　　要：

本文根据某组织搜集整理的 1998-2017 年全球发生的恐怖主义袭击事件记录信息，通过利用多种统计分析方法、数据分析软件，对大批量的数据进行处理和挖掘，主要完成了以下几个方面的工作：

针对问题一，本文提取了与恐怖袭击事件有相关性的人员伤亡、经济损失、袭击发生时机、攻击目标类型、攻击类型、使用武器类型等因素的记录信息，作为事件危害性的评级指标。首先对未定量的变量数据进行量化，然后对数据进行清洗处理，使数据标准化、无量纲并做归一化处理；其次，对数据做 KMO 和 Bartlett 检验，判断是否适合进行因子分析，检验通过后，使用 SPSS 软件计算出因子变量的得分，并计算出所有事件的综合得分及排名结果。同时，本文使用 CRITIC 赋权值法对指标变量进行赋值，增加了赋权的客观性，计算出所有事件的综合得分及排名结果，再对这两种方法所计算的综合得分进行标准化处理并求取均值，得到一组新的得分评价标准，然后对其进行从高到低排序处理，筛选出近二十年危害程度最大的 10 件恐怖袭击事件，最后采用 DBCLASD 聚类方法将标准化处理后的数据按危害程度分为了 1-5 级，并将任务 1 给出的事件划分等级。

对于问题二，针对在 2015、2016 年度发生的、尚未有组织或个人宣称负责的恐怖袭击事件，我们首先结合现实情况，仔细斟酌、选出衡量恐怖组织或个人的特征(指标)，再利用现有数据大致推断 2015 年和 2016 年的新生或隐藏的恐怖组织或个人的数量，利用数理统计软件 R 对这 2 年发生的未知恐怖组织或个人的样本进行聚类，确定新生的恐怖组织或个人的数量，并将他们作为恐怖事件的嫌疑人进行标记代号，接着求出每个嫌疑人的特征值，并根据问题一建立的模型求解每个嫌疑人的危害程度进行排序处理，将前 5 名嫌疑人标记为 1-5 号，最后确定题目所给表 2 恐袭事件的特征值，求出他们与所有嫌疑人相似性系数，然后按相似性系数从大到小排序，根据 1-5 号嫌疑人的位置，确定嫌疑性的大小，完善题目所给的表格。

问题三中，考虑到国际社会对恐怖袭击的伤亡人数关注度高，我们以伤亡人数数据为基础，从时空特性、蔓延特性、级别分布、主要原因四方面入手研究近三年来所发生的恐怖袭击事件。针对时空特性方面，先从 12 类地区按不同年份分析，找出其中伤亡比较严重的地区，再分析重灾区内各国家的伤亡人数情况，找出伤亡比较严重的国家，然后再对这些国家按月份分析；针对蔓延特性方面，也是根据伤亡人数对比分析 2015-2017 年不同地区热力分布图总结出恐怖事件发生的蔓延性；针对级别分类方面，我们运用 R 软件将问题 1 划分好的 5 类等级，按照经纬度坐标值画出全球分布矢量图，更直观地观察出恐怖事件在全球分布的状况。针对主要原因方面，我们结合附件 1 信息和互联网资料，总结出恐怖袭击发生的主要原因。通过这四个方面的细致分析，我们从中找寻恐怖事件发生的相关规律，分析预测了下一年全球或某些重点地区的反恐态势，并提出对反恐斗争的相关见解和建议。

对于问题四，利用附件 1 中的数据，使用商业销售领域常用的客户流失预测模型原理，建立恐怖组织“撤离预测模型”，以袭击月份（imonth）、袭击地区（region）和国家（country）、袭击目标类型（targtype1）、袭击类型（attacktypel）、使用武器类型（weapontypel）等作为评价指标，并筛选出数据并进行数据标准化、无纲量化、归一化处理，然后利用改进粒子群算法与支持向量机相结合的客户流失预测方法（IPSO-SVM），对数据集进行训练，最后根据 IPSO-SVM 的预测结果分析，分析出评价指标的影响程度，当地政府制定出针对这些指标变化的相应对策，即可对该地区的恐怖组织“撤离率”的大小进行预测和控制。

关键词：恐怖袭击事件，指标量化，因子分析，CRITIC 赋权，Pearson 相关系数分析，IPSO-SVM 支持向量机

目 录

一、问题重述

1.1 背景介绍

进入 21 世纪以来，恐怖主义已迅速蔓延至全球范围。据相关资料表明，自 2002 年起，恐怖袭击活动总体呈现愈演愈烈的趋势。我们时常会听到骇人听闻的恐怖袭击事件新闻报道，有发生在国外的，如英国伦敦今年多次遭受暴力袭击，偶尔国内也未幸免，如过去的“昆明火车站 301 暴恐”。恐怖主义袭击给当地人们带来了越来越严重的后果，伴随着人身侵害和财产物资损失导致人心惶恐，不仅一定程度上扰乱、破坏社会日常稳定的秩序，还阻碍经济增速发展，更重要的是将会持续影响到全人类的生存环境[1-2]。

恐怖主义已成为人类公敌，威胁着国际社会的和平与安全，打击恐怖主义活动是世界各国义不容辞的责任。如何做好反恐预警的准备并将恐怖主义活动及时消灭在谋划阶段成为越来越多国家与政府关注的焦点。而从海量信息中寻找到具有反恐情报预警价值的信息便是要开展的首要任务。深入分析恐怖袭击事件相关数据有利于加强人们对恐怖主义的认识，为反恐防恐提供有价值的信息支持[3]。

表 1-1 2002-2015 年各地区恐怖主义袭击发生次数[4]

地区	2002-2015 年发生恐怖袭击的次数	地区	2002-2015 年发生恐怖袭击的次数
北美	377	中亚	164
中美洲及加勒比地区	60	西欧	2116
南美	1792	东欧	3276
东亚	143	中东及北非	30406
东南亚	6999	撒哈拉以南的非洲地区	8577
南亚	29342	澳大利亚西区及大洋洲	36

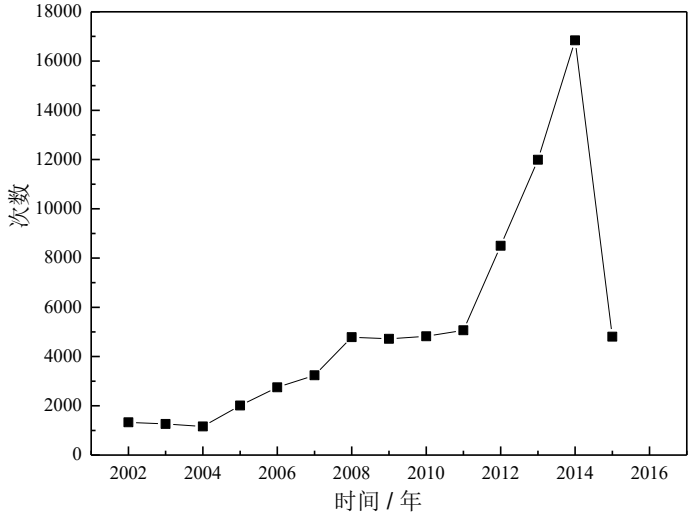


图 1.1 2002-2015 年全球恐怖袭击事件数量的发展趋势图[4]

1.2 问题提出

互联网与大数据时代已经到来，信息规模庞大，数据类别多样，非结构化数据使数据的处理变得更加棘手，信息杂声将有价值的信息淹没在数据海洋之中。由于信息的来源极为广泛，内容也颇为复杂，如何将信息去除杂质，仅仅搜集能高效生产情报的信息，但又不遗漏具有潜在价值的信息，保证达到信息源的精准高效运用。关键还需要设法构建精准的恐怖主义袭击风险评估模型与预测模型，为完善合理高效的情报反恐体系提供有力支撑[5]。

1.3 本文所要解决的问题

（1）恐怖袭击事件分级：根据影响危害性的诸多因素特征，基于数据分析的量化分级模型，采用因子分析法与 CRITIC 赋权值计算出综合得分，最后采用 DBCLASD 聚类方法将标准化处理后的数据按危害程度分为了 1-5 级，并将任务 1 给出的事件划分等级。

（2）恐怖袭击分子的侦查：结合已发生的但未有任何组织或个人宣称发起的恐怖袭击事件，提取相关特征，求解 2015-2016 年发生的未知恐怖组织或个人的样本与已知恐怖组织或个人样本的相似性，根据相似系数的大小判断未知恐怖组织或个人是新生的还是已有的，然后针对新生的恐怖组织或个人，我们根据问题一建立的模型所求得的危害性大小进行排序和标号，选出 1-5 号恐怖组织或个人，最后根据 1-5 号恐怖组织或个人与题目表 2 的恐袭事件的相似性，对 5 个恐怖组织或个人的嫌疑性进行排序，获取侦查结果。

（3）对未来反恐态势的分析：考虑到国际社会对恐怖袭击的伤亡人数关注度高，我们以伤亡人数数据为基础，从时空特性、蔓延特性、级别分布、主要原因四方面入手研究近三年来所发生的恐怖袭击事件。通过这四个方面的细致分析，我们从中找寻恐怖事件发生的相关规律，分析研判了下一年全球或某些重点地区的反恐态势，并从四个方面提出我们对反恐斗争的相关见解和建议。

（4）数据的进一步利用：利用附件 1 中的数据，依据商业销售领域常用的客户流失预测模型原理，建立恐怖组织“客户流失预测模型”，然后运用改进粒子群算法与支持向量机相结合的客户流失预测方法（IPSO-SVM），对数据集进行训练，分析出评价指标的影响程度。根据当地政府制定出针对这些指标变化的相应对策，即可对该地区的恐怖组织“撤离率”的大小进行预测和控制。

二、模型假设与符号说明

2.1 模型假设

a.假设遭受恐怖袭击地区与国家的人口密度、经济状况是平面内的均匀分布；

b.恐怖袭击产生的危害以人员伤亡人数来衡量，开始时的危害取决于人员伤亡与财产损失的多少；后续危害则包括：伤亡者处理费用、重新建设费用、原有生产模式的改变对经济的影响等等；

c. 将同一恐怖组织或个人在不同时间、不同地点多次作案的多个案件归为一类事件；

d. 每年新生或隐藏的恐怖组织或个人数量相差不大；

e. 恐怖袭击造成危害的消失与扩散因素：时间、距离相互独立(年月份与区域因素相互无影响)。

2.2 符号说明

符号	意义说明
X_1	人员伤亡总数，包括受害者与凶者在内的并已证实的数量
X_2	财产损失程度，由 4 个类别来描述。
X_3	袭击月份，记录了发生恐怖袭击事件的月份
X_4	地区国家，确定恐怖袭击事件发生的地区或国家
X_5	袭击目标类型，记录受害对象的一般类型，有 22 个类别
X_6	袭击类型，袭击的一般方法，由 9 个类别组成
X_7	武器类型，包含生物武器、化学武器、轻攻击武器等 13 类

三、问题分析

3.1 问题一的分析

问题一要求依据给出的附件 1 以及其它有关信息，结合现代信息处理技术，借助数学建模方法建立基于数据分析的量化分级模型，将附件 1 给出的事件按危害程度从高到低分为一至五级，列出近二十年来危害程度最高的十大恐怖袭击事件，并给出表 1 中事件的分级。

结合问题信息，本文提取了与恐怖袭击事件有相关性的人员伤亡、经济损失、袭击发生时机、攻击目标类型、攻击类型、使用武器类型等因素的记录信息，作为事件危害性的评级指标。首先对未定量的变量数据进行量化，然后对数据进行清洗处理，使数据标准化、无量纲并做归一化处理；其次，对数据做 KMO 和 Bartlett 检验，判断是否适合进行因子分析，检验通过后，使用 SPSS 软件计算出因子变量的得分，计算出所有事件的综合得分及排名结果。同时，本文使用 CRITIC 赋权值法对指标变量进行赋值，增加了赋权的客观性，计算出所有事件的综合得分及排名结果，再对这两种方法所计算的综合得分进行标准化处理并求取均值，得到一组新的得分评价标准，然后对其进行从高到低排序处理，筛选出近二十年危害程度最大的 10 件恐怖袭击事件，最后采用 DBCLASD 聚类方法将标准化处理后的数据按危害程度分为了 1-5 级，并将任务 1 给出的事件划分等级。

3.2 问题二的分析

问题二要求在于问题一获得的模型参数基础上，根据已知恐怖组织或个人的

样本求出每个恐怖组织或个人的特征值（用平均值表示定量特征、有序型特征的值，用众数表示无序型特征的值），再去求解 2015-2016 年发生的未知恐怖组织或个人的样本与已知恐怖组织或个人样本的相似性，根据相似系数的大小判断未知恐怖组织或个人是新生的还是已有的，然后针对新生的恐怖组织或个人，我们根据问题一建立的模型所求得的危害性大小进行排序和标号，选出 1-5 号恐怖组织或个人，最后根据 1-5 号恐怖组织或个人与题目表 2 的恐袭事件的相似性，对 5 个恐怖组织或个人的嫌疑性进行排序，并填充表格。

3.3 问题三的分析

问题三要求依据附件 1 并结合因特网上的有关信息，建立适当的数学模型，研究近三年来恐怖袭击事件发生的主要原因、时空特性、蔓延特性、级别分布等规律，进而分析研判下一年全球或某些重点地区的反恐态势，通过用图/表给出研究结果，提出对反恐斗争的一些见解和建议。

考虑到国际社会对恐怖袭击的伤亡人数关注度高，我们以伤亡人数数据为基础，从时空特性、蔓延特性、级别分布、主要原因四方面入手研究近三年来所发生的恐怖袭击事件。通过这四个方面的细致分析，我们从中找寻恐怖事件发生的相关规律，分析研判了下一年全球或某些重点地区的反恐态势，并从四个方面提出我们对反恐斗争的相关见解和建议。

3.4 问题四的分析

针对问题四：发挥题目附件 1 所给的数据的其他作用，为了更好地与恐怖组织作斗争，本文就恐怖主义的预防和控制展开研究，以过去近二十件的恐怖袭击事件为参考数据，借鉴了目前在电子商务领域，电信网络等比较依赖用户基数的商业领域，为了解决潜在的客户流失问题，常采用的基于统计学的方法建立客户流失状态模型。建立恐怖组织“撤离预测模型”，对潜在的恐怖组织状态进行判断。采用一种改进粒子群算法与支持向量机相结合的客户流失预测方法(IPSO-SVM)，对某地区历史恐怖袭击事件记录样本数据进行计算，得出分析结果后，当地政府针对某一个或者几个指标做出相应的应对政策，以提高恐怖组织在该地区的“撤离率”，达到对该地区遭受恐怖袭击事件预防和控制的效果。

四、模型建立与求解

4.1 问题一：依据危害性对恐怖袭击事件分级

4.1.1 影响危害性的因素选择与量化

袭击事件的危害性不仅取决于人员伤亡和经济损失这两个方面，还与发生的时机、地域、针对的对象等等诸多因素有关。其中人员伤亡与经济损失作为直接影响因素，通常与恐怖分子利用的武器类型、攻击类型、袭击目标有关。此外，恐怖袭击具有明显的地理空间属性，还与社会民主程度、所处政治角色有关，所以实施的对象也有差异。因考虑到原数据存在很多文本信息，对数据处理带来不方便，因此我们需要对这些因素进行量化分析处理[6]，这里选取了 7 个因素分析，但其中人数因素已给出数据，不需要量化，其他因素量化处理如下：

（1）财产损失程度对应权数

财产指标的衡量统一按照财产损失程度，包含四种程度：1 =灾难性的（可能大于 10 亿美元）、2 =重大的（可能大于 100 万美元，但小于 10 亿美元）、3 =较小的（可能小于 100 万美元）、4 =未知。根据财产损失的价值(美元)求解出每个等级对应的平均损失（不计空格），其中等级 1 没有对应的损失价值，取值为 10 亿。然后设等级 1 的权数为 1，由平均损失的比值关系确定其他等级的权数，其结果如表 4-1 所列。

表 4-1 财产损失等级对应权数

等级	1	2	3	4
金额/美元	1.00×10^{9}	13701322.63	137875.20	96000
权数	1	1.37×10^{-2}	1.38×10^{-4}	9.60×10^{-5}

（2）袭击月份权数量化处理

月份与伤亡人数水平的数值结果如表 4-2 所列。由此可见，每个月的袭击事件引起的伤亡人数分布规律不尽相同。五月、六月、七月、十月的因素权数较接近，也较其他月份偏大，说明在一定时间内发生的袭击事件次数下降较快，即这些月份与高伤亡恐怖袭击事件关联度较低。另外，二月、十二月的指标权数较小，说明在这两个月发生的恐怖袭击导致高伤亡人数的比例高于其他月份。伤亡人数最多的前十名恐怖袭击事件并没有发生在这段事件。这可能与恐怖主义组织选择袭击的实际具有偏向性有关，也可能与不同国家、不同民族具有重要意义节日的月份有关，这里参考的数据有限，需要进一步研究。

表 4-2 袭击月份因素数值结果

编号	1	2	3	4	5	6
权数	0.080	0.074	0.082	0.083	0.088	0.089
编号	7	8	9	10	11	12
权数	0.091	0.087	0.080	0.088	0.086	0.072

（3）地区与国家类型权数量化处理

由于恐怖袭击具有很强的地域性，地区与国家因素明确了恐怖袭击事件发生的区域。地区分为 12 类，并依赖于所提供资料中的国家编码：1 =北美、2 =中美洲和加勒比海地区、3 =南美、4 =东亚、5 =东南亚、6 =南亚、7 =中亚、8 =西欧、9 =东欧、10 =中东和东非、11 =撒哈拉以南的非洲、12 =澳大利亚和大洋洲。表 4-3 为地区与国家因素对应分析结果，可以看出，南亚、中东和北非的伤亡人数最多，而澳洲和中亚恐怖袭击引起的伤亡人数较少。从因素权数看出，东亚、中亚、东欧权数非常小，约 0.002～0.005。说明在这三个区域一旦发生恐怖袭击事件，极有可能导致大规模伤亡，2016 年 11.12 巴基斯坦重大爆炸袭击事件印证了该研究结果。这与当地社会经济和政府政治因素有关。北美作为全球综合实力最发达区域之一，政府投入了大量财政和军事力量应对恐怖主义，因此北美的总体恐怖袭击事件并不多，而且大部分恐怖袭击事件没有人员伤亡。但高伤亡恐怖

事件更倾向于发生在经济水平发达的地区和国家。因为在这些国家，恐怖组织更容易获取先进的武器和装备，从而发生破坏性的恐怖袭击[6]。

表 4-3 地区因素数值结果

编号	1	2	3	4	5	6
权数	0.009	0.085	0.135	0.002	0.073	0.255
编号	7	8	9	10	11	12
权数	0.003	0.083	0.005	0.244	0.083	0.021

（4）袭击目标类型权数量化处理

袭击目标类型因素变量包括以下 22 个类别：1 =商业、2 =政府、3 =警察、4 =军事、5 =流产有关、6 =机场和飞机、7 =政府（外交）、8 =教育机构、9 =食物或水供应、10 =新闻记者、11 =海事、12 =非政府组织(NGO)、13 =其他、14 =公民自身和私有财产、15 =宗教人物/机构、16 =电信、17 =恐怖分子/非州立民兵组织、18 =游客、19 =运输（航空除外）、20 =未知、21 =公用事业、22 =暴力政党。对于恐怖袭击，最重要的目标是要保证人员生命安全。只有了解恐怖组织更偏好袭击哪一类目标，以及哪一类目标容易出现高死亡，决策者才能制定可行而且有效的反恐策略。袭击目标与其后果的对应分析结果见表 4-4。由表可知，普通公民和个人财产是恐怖袭击频率和产生伤亡最多的目标，原因在于平民面对恐怖袭击通常没有准备以及缺乏自我保护。紧接着是警卫和军事目标，尽管这类组织具有较强的保护与抵抗能力，但恐怖分子为了实现政治、宗教或制造恐怖氛围的目的，更倾向于破坏警卫与军事目标。例如，武装集团为实现自己的利益在沦陷后的伊拉克的首要任务就是把美军驱赶出境内。大规模的有计划、有预谋针对军事目标的恐怖袭击可能是导致军事目标高死亡的原因。

表 4-4 袭击目标因素数值结果

编号	1	2	3	4	5	6	7	8	9	10	11
权数	0.075	0.128	0.168	0.163	0.001	0.004	0.012	0.014	0.001	0.011	0.002
编号	12	13	14	15	16	17	18	19	20	21	22
权数	0.005	0.015	0.284	0.023	0.04	0.019	0.003	0.037	0.008	0.004	0.013

（5）袭击类型权数量化处理

袭击类型即攻击的一般方法，由 9 个类别组成，分别定义为：1 =暗杀、2 =劫持、3 =绑架、4 =路障事件、5 =轰炸/爆炸、6 =未知、7 =武装突袭、8 =徒手攻击、9 =设施/基础设施攻击。基于恐怖袭击分子所采用的袭击方法频率以及该方法所造成的人员伤害数量，结合各种调查数据综合分析，所获得袭击方式结果见表 4-5 所示。由表可知，不同袭击方法结果差异较大，表明不同袭击方法造成的死亡人数分布规律差异性更强。暗杀、绑架及基础设施攻击具有近似轮廓，路障事件、徒手攻击具备近似轮廓。劫持、绑架易造成高伤亡事件，多次劫机事件造成的后果就说明了这一点。暗杀一般针对的为特定个体，不会造成大范围人群伤亡，徒手攻击引起的伤亡概率最小。

表 4-5 袭击类型因素数值结果

编号	1	2	3	4	5	6	7	8	9
权数	0.263	0.388	0.273	0.002	0.003	0.032	0.010	0.002	0.027

（6）武器类型权数量化处理

恐怖袭击运用的武器被分为 13 类：1 =生物武器、2 =化学武器、3 =放射性武器、4 =核武器、5 =轻武器、6 =爆炸物/炸弹/炸药、7 =假武器 、8 =燃烧武器 、9 =致乱武器、10 =交通工具、11 =破坏设备、12 =其他、13 =未知。我们通过 Excel 筛选出每类武器使用后的死亡人数、受伤人数和使用次数，总伤害的数值是通过赋给死亡人数和受伤人数相应的权重计算出来的，即总伤害= 1×死亡人数 + 0.5 ×受伤人数，平均伤害=总伤害/次数，这里最后是以平均伤害的数值来衡量武器的危害性，即武器类型的权数，其分析结果如表 4-6 所示。由表可知，不同武器产生的结果差异较大。其中利用车辆等交通工具、化学武器发动的袭击易造成较多伤亡人数，接着为轻武器、爆炸物类。轻武器可造成人员直接伤亡的结果，而化学武器具有持续危害性作用。随着恐怖分子愈来愈有规划，寻得生化武器的概率大大增加。因此，各个国家安全职能部门应加强关注交通工具和化学武器，并严格控制这些武器的流通渠道。

表 4-6 武器类型因素数值结果

编号	1	2	3	4	5	6	7
死亡	10	510	2	0	89450	149630	1
受伤	27	6656	4	0	46775	313992	0
总的伤害	23.5	3838	4	0	112837.5	306626	1
次数	27	230	11	0	33316	64045	12
平均伤害	0.8704	16.6870	0.3636	-	3.3869	4.7877	0.0833
编号	8	9	10	11	12	13	
死亡	3787	5076	3126	52	107	27795	
受伤	3435	2915	17038	158	64	10502	
总的伤害	5504.5	6533.5	11645	131	139	33046	
次数	5436	2234	108	106	84	8574	
平均伤害	1.0126	2.9246	107.8241	1.2358	1.6548	3.8542	

4.1.2 模型建立与求解

（1）模型的构建与方法分析

因子分析是一种多元统计分析方法，核心思想是数据变换与降维，先把错综复杂的变量综合成少数主要因子，再进行问题解释或综合评价[8]，它可用少量潜在因子解释原始变量大部分信息[9]。因子分析的出发点是原始变量的相关矩阵。

因子分析可消除变量间的相关性，通过把恐怖袭击风险指标进行数学变换，综合成几个因子，根据一定标准选取主因子，得到每个恐怖袭击事件危害程度综合得分；同时因子不需要主观确定指标权重，而是根据样本数据的观测值自动得到权重，因此可以消除主观因素，提供客观的评价结果。

本文拟选取人员伤亡总数 X_1、财产损失程度 X_2、袭击月份 X_3、受袭地区国家 X_4、袭击目标类型 X_5、袭击类型 X_6、武器类型 X_7 作为 7 个因素指标。这 7 个变量标准化后，变量的平均值为 0，标准差为 1，每个变量又可由 $k(k<p)$ 个因子线性组合来表示。则数学模型可建立如下：

$$\begin{cases} x_1 = \alpha_{11}f_1 + \alpha_{12}f_2 + \alpha_{13}f_3 + \cdots \alpha_{1k}f_k + \theta_1 \\ x_2 = \alpha_{21}f_1 + \alpha_{22}f_2 + \alpha_{23}f_3 + \cdots \alpha_{2k}f_k + \theta_2 \\ \cdots \\ x_p = \alpha_{p1}f_1 + \alpha_{p2}f_2 + \alpha_{p3}f_3 + \cdots \alpha_{pk}f_k + \theta_p \end{cases} \tag{4-1}$$

该模型也可以用矩阵形式表示：

$$x = \alpha f + \theta \tag{4-2}$$

其中：

$$\alpha = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1k} \\ \vdots & \ddots & \vdots \\ \alpha_{p1} & \cdots & \alpha_{pk} \end{bmatrix} \tag{4-3}$$

其中，f 表示公共因子；x 为标准化后的原始变量；α 为因子载荷矩阵，而且 $\alpha_{\mathrm{ij}}(i = 1,2,3 \ldots, k.j = 1,2,3 \ldots p)$；$\alpha_{\mathrm{ij}}$是：$f_i$和$x_i$的协方差，其中$\alpha_{\mathrm{ij}}$绝对值越大意味着$x_i$和$f_i$的依赖程度越大，反之也一样。

本文采用主成分分析的方法求解因子载荷矩阵，具体的计算过程如下：

① 数据的标准化

正规化原始变量数据采用如下方法：

$$y_i = \frac{x_i - \bar{x}}{s}, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4-4}$$

则新的矩阵$\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1k} \\ \vdots & \ddots & \vdots \\ \alpha_{p1} & \cdots & \alpha_{pk} \end{bmatrix}$元素的均值为 0，方差为 1，且为无量纲。

② 计算原始样本协方差矩阵

设$\mathrm{X} = (X_1, X_1, X_1 \ldots X_7)$为 7×N 的矩阵，N 代表本文筛选后样本数，则矩阵：

$$\mathrm{C} = \left(c_{i,j}\right)_{7\times N} = \begin{pmatrix} c_{11} & \cdots & c_{17} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{n7} \end{pmatrix} \tag{4-5}$$

其中，$C_{i,j} = \mathrm{Cov}$（X_i，X_j），$i, j = 1,2,3\cdots n$，C_{ij} 为 X 的X_i，X_j的协方差。

③ 计算协方差矩阵非零特征根

利用特征方程$(\lambda E - A)x = 0$计算协方差矩阵的非零特征根并排序：$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, 相应的单位正交化特征向量$e_i(i = 1,2,3 \ldots n)$。

④ 计算因子载荷矩阵

$$A = (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \sqrt{\lambda_3}e_3, \ldots \sqrt{\lambda_n}e_n) \quad (4\text{-}6)$$

⑤ 因子旋转

本文采用正交旋转最大方差法对初始公因子做线性组合，即因子旋转，使综合因子具有特定含义[10]。最大方差法通过旋转把公共因子变量载荷的方差最大化，从而使因子上某些变量荷载系数朝着最大或最小方向变化，保证中等大小载荷没有或很少，从而使得因子的意义更具体，便于命名因子。解决了原始变量综合为少数几个因子后，若因子含义模糊，不利于进一步解释与评价的问题[11]。因子旋转可保证新因子更接近零或远离零，α_{ij} 载荷接近零，说明公因子 x_i 和 f_j 相关性弱，载荷接近绝对值 1 说明相关性强。因此，经过因子旋转，共同因子的实际意义更加明确。

（2）模型的求解

本文这里借助 SPSS 软件工具完成因子分析的具体步骤如下：

① 选取原始变量，标准化处理；

② 求解变量相关系数矩阵；

③ KMO 检验和巴特利球形检验，确定因子分析的适用性；

④ 求解初始公共因子及因子载荷矩阵；

⑤ 因子旋转，对主因子命名；

⑥ 求得因子得分系数矩阵，计算因子得分；

⑦ 计算综合得分，进行综合评价。

原始变量的选取、标准化及求解变量相关系数矩阵的工作前文已经做过，下面对原始数据进行检验，判断原始数据是否适合因子分析，本文使用 KMO 检验（Kaiser-Meyer-Olkin）和巴特利球形检验，

a. KMO 检验

KMO 检验统计量用来对比变量间的相关系数和偏相关系数，统计量为：

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}^2}{\sum\sum_{i \neq j} r_{ij}^2 + \sum\sum_{i \neq j} p_{ij}^2} \quad (4\text{-}7)$$

式中，r_{ij}为变量X_i与变量X_j的简单相关系数；p_{ij}为X_i与X_j的偏相关系数。统计量取值范围[0，1]。KMO 越接近 1，说明变量间相关性越强，越适合因子分析；KMO 接近 0，说明变量间相关性非常弱，不适合因子分析。大于 0.9 则表示非常适合；0.7~0.8 适合；0.5~0.7 不太适合；小于 0.5 不适合。

b. 巴特利球形检验

巴特利球形检验统计量由原始变量相关系数矩阵得到，近似服从卡方分布。假设相关系数矩阵为单位阵，如果卡方值显著且 P 小于 0.05, 拒绝原假设，说明变量间存在相关性，即原始变量适合因子分析。反之，相应大于 0.05，则接受原假设，认为相关系数矩阵可能为单位矩阵，不适合做因子分析。

表 4-7 KMO 和巴特利球形 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量	0.683
Bartlett 的球形度检验 近似卡方	60835.480
df	21
sig	0.000

表中的 KMO 取值为 0.683，表明可以进行因子分析；巴特利球形检验是为了看数据是否来自服从多元正态分布的总体，表中 Sig 值为 0.000，说明数据来自于正态分布总体，适合做进一步分析。

变量共同度表示数据中各变量中所含原始信息能被提取的公因子所解释的程度。如下表所示，本文选取的变量共同度都在 65%以上，所以提取的这几个公因子对变量的解释能力较强。

表 4-8 公因子方差

	初始	提取
Zscore（人员伤亡总数 X_1）	1.000	0.761
Zscore（财产损失程度 X_2）	1.000	0.790
Zscore（袭击月份 X_3）	1.000	0.654
Zscore（地区国家 X_4）	1.000	0.676
Zscore（袭击目标类型 X_5）	1.000	0.857
Zscore（袭击类型 X_6）	1.000	0.755
Zscore（武器类型 X_7）	1.000	0.654

提取方法：主成分分析

又如下表，解释的总方差“初始特征值”一栏只有前 3 个特征值大于 1，所以只提取前 3 个主成分；表中“提取平方和载入”前 3 列成分方差占所有主成分方差的 83.535%，由此可见，选前 3 个主成分足够代替原来的变量，几乎涵盖了原变量的全部信息，所占的权重分别是

$$\begin{cases}\omega_1 = 0.34187\\ \omega_2 = 0.25029\\ \omega_3 = 0.24309\end{cases}$$

“旋转平方和载入”一栏显示的是旋转以后的因子提取结果，与未旋转之前差别不大。

表 4-9 解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%	合计	方差的%	累积%
1	1.693	34.187	34.187	1.693	34.187	34.187	1.693	34.182	34.182
2	1.052	25.029	59.216	1.052	25.029	59.216	1.051	25.020	59.201
3	1.002	24.309	83.525	1.002	24.309	83.525	1.003	24.324	83.525
4	0.999	6.268	89.793						
5	0.972	3.891	93.684						
6	0.910	0.996	94.680						
7	0.372	0.320	100.000						

提取方法：主成分分析

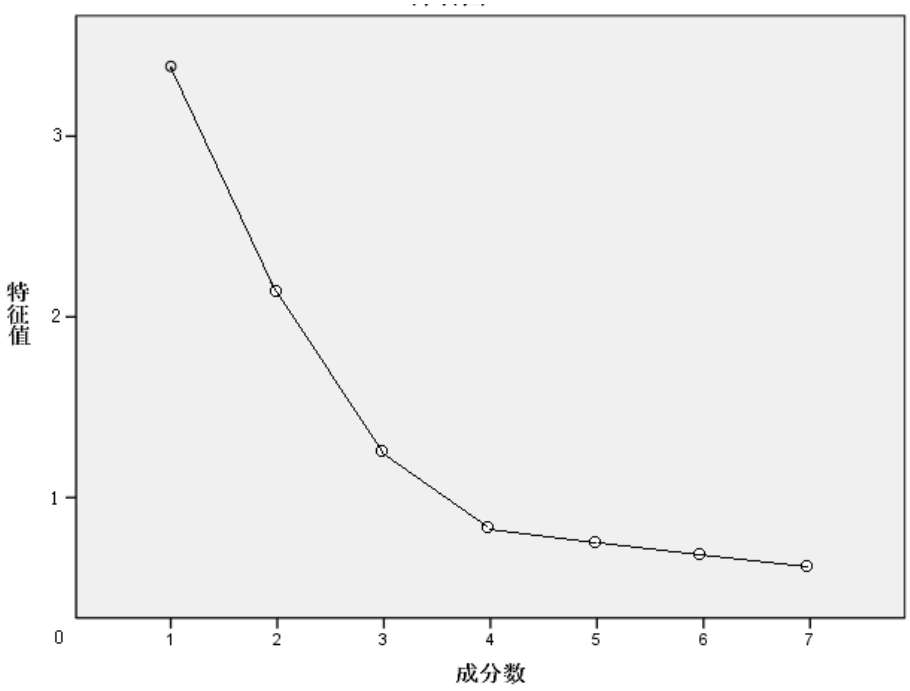


图 4.1 成分-特征曲线图

如图 4.1 所示，有 3 个成分的特征值超过了 1，所以可以只考虑这 3 个成分即可。由旋转成分矩阵表可以反映第一个成分因子在“人员伤亡人”、“武器类型”、“攻击类型”、“袭击目标类型”、“袭击月份”、“财产损失程度”和“国家地区”中占有较大的载荷，所以其反映其他变量的信息。

表 4-10 旋转成份矩阵

	成份		
	1	2	3
Zscore（人员伤亡总数 X_1）	0.871	-0.039	0.006
Zscore（财产损失程度 X_2）	0.886	-0.066	0.000
Zscore（袭击月份 X_3）	-0.001	-0.134	-0.369

Zscore（地区国家 X_4）	-0.006	0.758	-0.031
Zscore（袭击目标类型 X_5）	0.004	-0.109	0.919
Zscore（袭击类型 X_6）	-0.023	0.578	0.143
Zscore（武器类型 X_7）	0.384	0.326	-0.002

提取方法：主成分分析。

旋转法：具有 Kaiser 标准化的正交旋转法。旋转在 4 次迭代后收敛。

由成分得分系统矩阵，如下表所示：

表 4-11 成份得分系数矩阵

	成份		
	1	2	3
Zscore（人员伤亡总数 X_1）	0.515	-0.45	0.003
Zscore（财产损失程度 X_2）	0.524	-0.070	-0.002
Zscore（袭击月份 X_3）	0.002	-0.125	-0.368
Zscore（地区国家 X_4）	-0.010	0.722	-0.036
Zscore（袭击目标类型 X_5）	0.000	-0.110	0.917
Zscore（袭击类型 X_6）	-0.019	0.549	0.139
Zscore（武器类型 X_7）	0.224	0.307	-0.005

提取方法：主成份分析。

旋转法：具有 Kaiser 标准化的正交旋转法。构成得分。

可以得出各个公因子的表达式如下：

$$FAC\_1 = 0.515 \times \alpha_{ij} + 0.224 \times \alpha_{ij} - 0.019 \times \alpha_{ij} + 0.00 \times \alpha_{ij} + 0.002 \times \alpha_{ij} + 0.524 \times \alpha_{ij} - 0.010 \times \alpha_{ij} \quad (4-8)$$

$$FAC\_2 = -0.045 \times \alpha_{ij} + 0.307 \times \alpha_{ij} - 0.549 \times \alpha_{ij} - 0.110 \times \alpha_{ij} - 0.125 \times \alpha_{ij} - 0.070 \times \alpha_{ij} + 0.722 \times \alpha_{ij} \quad (4-9)$$

$$FAC\_3 = 0.003 \times \alpha_{ij} - 0.005 \times \alpha_{ij} + 0.139 \times \alpha_{ij} + 0.917 \times \alpha_{ij} - 0.368 \times \alpha_{ij} - 0.002 \times \alpha_{ij} - 0.036 \times \alpha_{ij} \quad (4-10)$$

其中 $(i, j) i = 1, 2, 3 \cdots n, j = 1, 2, 3 \cdots 7$。

即可计算出每一个样本事件的综合得分W_i

$$W_i = \text{FAC\_1}_i \times \omega_1 + \text{FAC\_2}_i \times \omega_2 + \text{FAC\_3}_i \times \omega_3 \qquad (4-11)$$

采用因子分析法按危害程度的综合得分降序排列如下：

事件编号	FAC1_1	FAC2_1	FAC3_1	综合得分
200109110004	201.12796	-14.27307	0.56717	46.58
200109110005	201.12796	-14.27307	0.56717	46.58
200109110006	97.52514	-5.23923	-0.53163	22.72
200109110007	94.77832	-5.05104	-0.13352	22.15
199808070002	48.33829	-5.09733	-0.25348	10.89
201603080001	17.53945	0.18858	0.51477	4.34
200102140003	7.06325	10.66137	0.59572	3.40
200806190010	1.28318	-2.06369	23.35000	3.34
200807020004	1.04931	-2.08781	23.21761	3.26
200102100006	-0.05146	-1.42346	24.33174	3.26
201202100020	-0.11851	-1.55328	24.33412	3.22
201212060017	-0.03026	-1.91543	24.32713	3.19
200612070003	-0.03384	-1.91495	24.32714	3.18
201112170024	-0.03384	-1.91495	24.32714	3.18
201612300002	-0.03384	-1.91495	24.32714	3.18
199908100003	6.85435	10.39146	-0.25606	3.18
201012020006	0.00450	-2.01306	24.33209	3.18
201309200028	-0.05797	-1.55569	23.93923	3.18
200912110015	-0.01783	-2.01112	24.33195	3.18
201002170024	0.01145	-1.96328	24.19659	3.17
200512120004	-0.04097	-2.00855	24.33191	3.17
201312220009	-0.04097	-2.00855	24.33191	3.17

图 4.2 运用因子分析法获取的部分数据截图

（3）模型的优化

虽然因子分析法具有化繁为简的优点，然而它不是对原有变量的取舍，而是根据原始变量的信息进行重新组合，找出影响变量的共同因子，化简数据，并通过旋转使得因子变量更具有可解释性，命名清晰性高。但是，在计算因子得分时，采用的是最小二乘法，此法有时可能会失效，是结果产生偏差。

本文采用 CRITIC 赋值法对模型进行优化。Critic 赋值法以两个基本概念为基础：一是对比强度，借鉴标准离差法的思想，认为若同一指标的所有评价指数差别越大，即标准差越大，则所蕴含的信息量越大；二是评价指标之间的冲突性，指标之间的冲突性是以指标之间的相关系数为基础，如两个指标之间具有较强的正相关，说明两个指标冲突性较低。第 j 个指标与其它指标的冲突性的量化指标 $\sum_{i=1}^{n}(1-r_{ij})$，其中r_{ij}评价指标i和 j 之间的相关系数。各个指标的客观权重确定就是对比强度和冲突性来综合衡量的。设C_j表示第 j 个评价指标所包告的信息量。C_j的计算式：

$$C_j = \delta_j \sum_{i=1}^{n}(1-r_{ij}) \qquad (4\text{-}12)$$

其中，n 为同一指标的评价数量。

一般地， C_j越大，第 j 个评价指标所包含的信息量越大，则该指标的相对重要性也就越大。设W_j为第 j 个指标的客观权重。

W_j的计算公式：

$$W_j = \frac{c_j}{\sum_j^m c_j} \qquad (4\text{-}13)$$

其中， m 为所有标量的数量。

使用 CRITIC 赋值法得到的综合得分及按照恐怖袭击的危害程度降序排列的结果如下：

	C1	C2	C3	C4	C5	C6	C7	
	袭击月份	人员伤亡总数	财产损失程度	地区国家	袭击目标类型	攻击类型	武器类型	
袭击月份	1	-4.78127E-05	-0.003578337	-0.010126554	-0.002171438	-0.001969472	-0.003715969	
人员伤亡总数	-4.78127E-05	1	0.624347357	0.008806013	0.003987807	0.006535163	0.129535463	
财产损失程度	-0.003578337	0.624347357	1	-0.019356327	0.001502655	-0.016469781	0.178687542	
地区国家	-0.010126554	0.008806013	-0.019356327	1	-0.00382635	0.036204385	0.04603379	
袭击目标类型	-0.002171438	0.003987807	0.001502655	-0.00382635	1	0.005232691	0.002776276	
攻击类型	-0.001969472	0.006535163	-0.016469781	0.036204385	0.005232691	1	0.007377327	
武器类型	-0.003715969	0.129535463	0.178687542	0.04603379	0.002776276	0.007377327	1	
	0	1.000047813	1.003578337	1.010126554	1.002171438	1.001969472	1.003715969	
	1.000047813	0	0.375652643	0.991193987	0.996012193	0.993464837	0.870464537	
	1.003578337	0.375652643	0	1.019356327	0.998497345	1.016469781	0.821312458	
	1.010126554	0.991193987	1.019356327	0	1.00382635	0.963795615	0.95396621	
	1.002171438	0.996012193	0.998497345	1.00382635	0	0.994767309	0.997223724	
	1.001969472	0.993464837	1.016469781	0.963795615	0.994767309	0	0.992622673	
	1.003715969	0.870464537	0.821312458	0.95396621	0.997223724	0.992622673	0	
sum	6.021609584	5.22683601	5.234866891	5.942265043	5.992498359	5.963089688	5.639305571	
std	0.005619074	46.12760024	0.006001578	0.083809392	0.343857903	0.117822563	3.376460221	
Cj	0.033835869	241.101402	0.03141746	0.498017623	2.060567917	0.702586511	19.04089094	263.4687
Wj	0.141913991	0.149453703	0.150985857	0.140325697	0.141228234	0.140653003	0.135439516	
	0.379338914	0.399492714	0.403588192	0.375093374	0.377505871	0.375968269	0.362032514	2.67302
Wj2	0.000128425	0.915104471	0.000119246	0.001890234	0.007820921	0.002666679	0.072270025	

图 4.3(a) 运用 CRITIC 赋值法获取的部分数据截图

事件编号	得分
200109110004	1445.68
200109110005	1445.68
199808070002	632.01
201603080001	227.01
200802010006	174.07
200409010002	160.73
200607120001	150.94
200708150005	150.23
200708160008	150.23
201710010018	136.54
201710140002	135.83
200403210001	110.28
200509140001	105.69
201404150089	103.16
201406100042	100.91
200908190001	98.50
201606010046	97.22
201705310001	88.02
201607020002	87.88
201509280037	80.71
200403110003	78.88
201607140001	78.29

图 4.3(b) 运用 CRITIC 赋值法获得的综合得分部分数据截图

因子分析法和 CRITIC 赋值法分别得到的结果采用加权综合的方法，获取综

合的排序结果得到近二十年危害程度最大的 10 件恐怖袭击如图所示：

表 4-12 不同方法计算得出的 10 大恐怖事件

因子分析法	CRITIC 法	两种方法综合分析
200109110004	200109110004	200109110004
200109110005	200109110005	200109110005
200109110006	199808070002	199808070002
200109110007	201603080001	200109110006
199808070002	200802010006	200109110007
201603080001	200409010002	201603080001
200102140003	200607120001	200802010006
200806190010	200708150005	200708150005
200807020004	200708160008	200708160008
200102100006	201710010018	200607120001

本文使用基于密度的聚类方法（DBCLASD）对前面加权综合得到的恐怖袭击事件危害程度排序数据事件按危害程度从高到低分为一至五级。

其主要思想为:确定簇的核心思想是基于密度的方法，即与其它区域比，样本空间中的簇显得更稠密。稠密区域显著特征是域内点的最邻近距离要小于域外点的最邻近距离。最邻近距离的概率分布用于描述由点集所形成的簇的特征，也可用于测试一个邻近点是否应被包含于簇内。

样本空间被分解成 p−维单元，其中单元边长被设定为NNDSet(S)的最大元素，其中S为点集, NNDSet(S)为点集的最邻近距离之集。如果一个单元包含一个或更多点集，那么就称该单元被“占有”，集合S容量的近似值即被占有的所有格单元之和。倘若结果簇的NN DSet(S)仍拟合期望距离分布，DBCLASD算法增量地把邻近点加入到初始簇。使用MATLAB工具对数据分级处理的结果如下：

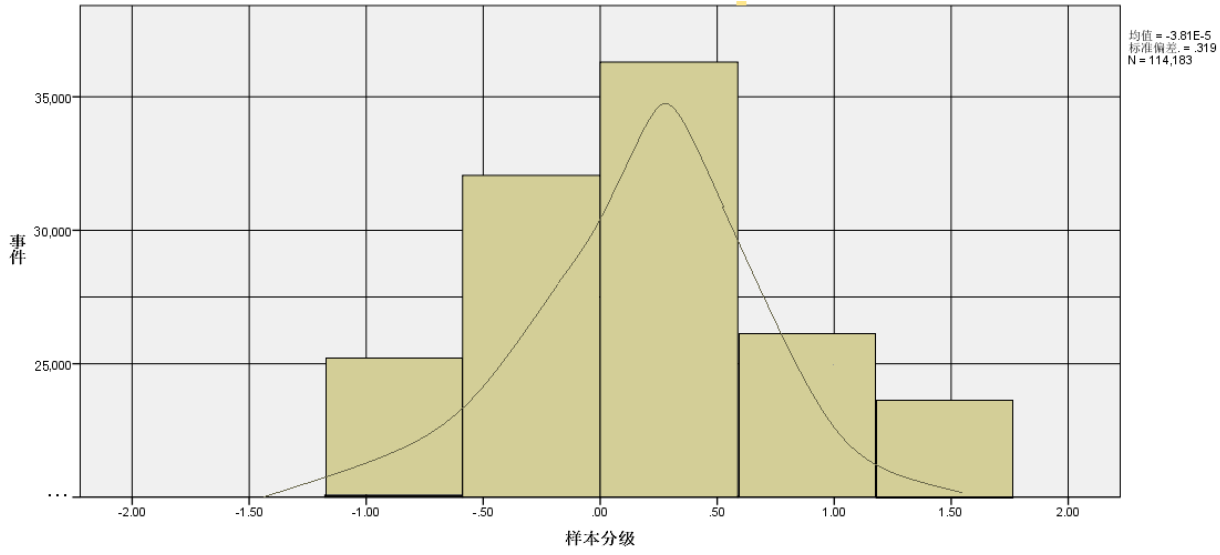


图 4.4 运用 MATLAB 工具对数据分级处理的结果

所给表 4-13 中事件的分级结果：

表 4-13 典型事件危害级别

事件编号	危害级别
200108110012	1
200511180002	1
200901170021	5
201402110015	3
201405010071	5
201411070002	2
201412160041	1
201508010015	5
201705080012	1

4.2 问题二：依据事件特征发现恐怖袭击事件制造者

4.2.1 分析与求解思路

根据对题目和数据的理解，确定问题二的解决思路如下：

（1）结合现实情况，仔细斟酌、选出衡量恐怖组织或个人的特征(指标)。

（2）利用现有数据大致推断 2015 年和 2016 年的新生或隐藏的恐怖组织或个人的数量，利用数理统计软件 R 对这 2 年发生的未知恐怖组织或个人的样本进行聚类，确定新生的恐怖组织或个人的数量，并将他们作为恐怖事件的嫌疑人进行标记代号。

（3）求出每个嫌疑人的特征值，并根据问题一建立的模型求解每个嫌疑人的危害程度，然后排序，将前面的 5 名嫌疑人标为 1-5 号嫌疑人。

（4）确定题目表 2 恐袭事件的特征值，求出它们与所有嫌疑人相似性，然后按相似性从大到小排序，根据 1-5 号嫌疑人的位置，确定嫌疑性的大小。

4.2.2 模型建立与求解

（1）模型的构建

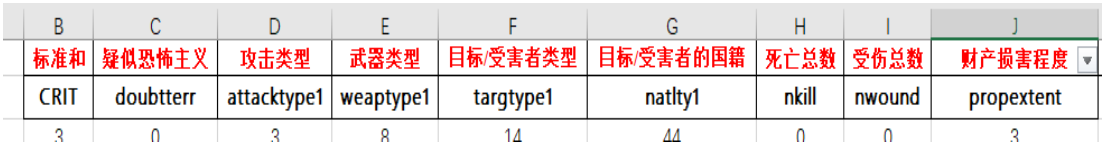
本问题的的解决涉及求解变量之间的相似性，而相似性可以从两方面考虑，一方面是计算距离，比如欧几里得距离（Eucledian Distance）、曼哈顿距离（Manhattan Distance）、明可夫斯基距离（Minkowski distance）等；另一方面是求解相似系数，比如 Jaccard 系数、余弦相似度（Cosine Similarity）、皮尔森相关系数(Pearson Correlation Coefficient)等。本文采用常用的皮尔森相关系数衡量样本之间的相似性[12]。

Pearson 相关系数又称相关相似性，通过 Peason 相关系数来度量两个用户的相似性[13]。计算时，首先找到两个用户共同评分过的项目集，然后计算这两个向量的相关系数。其计算公式如下：

$$\mathrm{r(X,Y)} = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} * \sqrt{n\sum y^2 - (\sum y)^2}} \tag{4-14}$$

（2）特征的提取

每一个恐怖组织或个人都是在特定的环境背景下产生的，具有自己的特点，而有些特点是共有的。根据对所给材料的理解，并结合实际情况，为了对恐怖组织或个人进行界定，筛选出相应的衡量指标包括标准和(CRIT)、疑似恐怖主义（doubtterr）、攻击类型(attacktype1)、武器类型(weaptype1)、目标/受害者类型(targtype1)、目标/受害者的国籍(natlty1)、死亡总数(nkill)、受伤总数(nwound)和财产损害程度(propextent)（注：附上相应英文字段在中文字段后），其中标准和（CRIT）为标准 1 至标准 3 综合而成 1 个字段，其值为三个字段之和。



B	C	D	E	F	G	H	I	J
标准和	疑似恐怖主义	攻击类型	武器类型	目标/受害者类型	目标/受害者的国籍	死亡总数	受伤总数	财产损害程度
CRIT	doubtterr	attacktype1	weaptype1	targtype1	natlty1	nkill	nwound	propextent
3	0	3	8	14	44	0	0	3

图 4.5 Excel 表格数据处理截图

根据附件表格，分别筛选出已确定恐怖组织或个人的样本（设为样本 Sk）和未确定恐怖组织或个人的样本(设为样本 Su)（详见附件），然后在 Su 中筛选出事件发生于 2015 年和 2016 年的样本 S_{56}。

在所有特征变量中，标准和、疑似恐怖主义、攻击类型、武器类型、目标/受害者类型和目标/受害者的国籍均为无序型变量，死亡总数、受伤总数为标量型变量，财产损害程度为有序型变量。对样本 Sk 进行统计得知，已知恐怖分子或组织有 1582 个，根据恐怖分子或组织分组并排除缺失值，然后无序型变量取众数，标量型变量与有序型变量取平均值。由此确定恐怖分子或组织的特征值（详见附件）。对于样本 S_{56} 的特征缺失值，无序型特征的用该列众数替代，标量型变量与有序型变量用该列平均值替代。

（3）相似度计算

将处理好的恐怖分子或组织的数据与样本 S_{56} 按列竖向汇合，用数据统计工具 R 中的函数 cor()计算 S_{56} 每个样本与 1500 多个恐怖分子或组织的相关系数，部分相关系数如下表 4-14 所示（详见附件）。

表 4-14 S_{56} 部分样本与部分恐怖分子或组织的相关系数

Sample NO.	14.K.Triad	14.March.Coalition	28s	313.Brigade.(Syria)
201602050060	0.975513	0.764499	0.999084	0.985024
201602050061	0.965189	0.776464	0.999898	0.986256
201602050063	0.964441	0.779570	0.999848	0.986871
201602050067	0.954827	0.774333	0.998642	0.982031
201602050069	0.989328	0.754286	0.993993	0.982587
201602050071	0.989328	0.754286	0.993993	0.982587

每个样本从 1500 多个相关系数选出最大值以及对应的恐怖分子或组织，并作图进行分析，如图 4.6 所示。

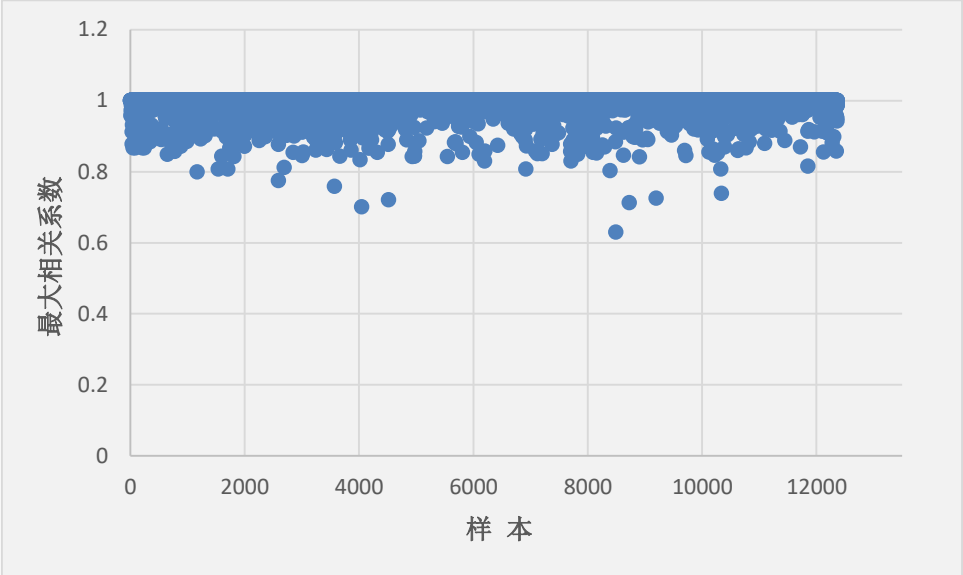


图 4.6 S_{56} 样本的最大相关系数

由上图可以看出，S_{56} 每个样本的最大相关系数普遍比较大，这是因为样本的特征值与已知恐怖分子或组织的普遍比较接近。

根据选取的最大相关系数进行排序，将值小于 0.9 的样本选出来作为样本 Sx，认为这些事件是新兴的恐怖分子或组织所为；值大于 0.9 的认为是现有的恐怖分子或组织所为。

在问题一中，我们已求得附件所有样本事件的危害程度得分，由此可知 Sx 中每个新兴的恐怖分子或组织的危害性大小。根据危害程度得分进行降序排序，给第 1 至第 5 个新兴恐怖分子或组织标号为嫌疑人 1 号至 5 号，如表 4-15 所列。

表 4-15 1-5 号嫌疑人及其特征

eventid	crit	doubtterr	attacktype1	weaptype1	targtype1	natlty1	nkill	nwound	propextent
1 号嫌疑人	3	0	3	6	4	4	34	23	3.0000
2 号嫌疑人	3	0	8	2	8	4	0	18	3.1593
3 号嫌疑人	2	1	2	5	4	4	4	8	3.1593
4 号嫌疑人	3	0	2	5	3	4	0	10	3.1593
5 号嫌疑人	3	0	3	6	3	4	3	10	3.0000

（4）嫌疑人排序

对于题目中的表 2 给出的 10 个未确定恐怖分子或组织的样本，设为 S10，首先找出它们的各个特征值，然后与 Sx 汇合，再次利用 R 求取 10 个样本与 Sx 每个新兴恐怖分子或组织的相关系数。对相关系数进行排序，根据 1 号至 5 号恐怖分子或组织的位置，填补表 2，如表 4-16 所列。

表 4-16 嫌疑人相关系数

样本编号	1 号嫌疑人	2 号嫌疑人	3 号嫌疑人	4 号嫌疑人	5 号嫌疑人
201701090031	12335	12349	12166	12256	12309
201702210037	8	107	20	92	19
201703120023	12356	12237	11905	11891	11941
201705050009	12359	12318	11935	11925	12088

201705050010	12359	12318	11935	11925	12090
201707010028	12324	12353	12295	12336	12329
201707020006	11291	12298	11893	11907	11939
201708110018	372	6497	723	12313	2675
201711010006	43	30	10	64	41
201712010003	12206	12355	12053	12292	12285

针对表 4-16 的每个恐怖袭击事件的样本编号按照每行的相关系数大小进行排序，然后结合问题一的模型求解，计算出所有嫌疑人的危害度，并对其进行排序，得到危害程度最大前五名嫌疑人：Islamic State of Iraq and the Levant (ISIL)、Taliban、Houthi extremists (Ansar Allah)、Kurdistan Workers' Party (PKK)、Palestinian Extremists，最后得出如表 4-17 中的恐怖分子关于典型事件的嫌疑度。

表 4-17 恐怖分子关于典型事件的嫌疑度

Sample NO.	Islamic State of Iraq and the Levant (ISIL)	Taliban	Houthi extremists (Ansar Allah)	Kurdistan Workers' Party (PKK)	Kurdistan Workers' Party (PKK)
201701090031	2	1	5	4	3
201702210037	5	1	3	2	4
201703120023	1	2	4	5	3
201705050009	1	2	4	5	3
201705050010	1	2	4	5	3
201707010028	4	1	5	2	3
201707020006	5	1	4	3	2
201708110018	5	2	4	1	3
201711010006	2	4	5	1	3
201712010003	4	1	5	2	3

4.3 问题三：对未来反恐态势的分析

4.3.1 现状导向

为了进一步提高未来反恐斗争的针对性和效率，结合 2015-2017 三年间的恐怖袭击相关信息，分别从恐怖袭击发生的时空特性、蔓延特性、级别分布和主要原因进行分析，然后结合这四方面的分析来对未来全球和重点地区的恐怖袭击趋势做相应的预测。

由于恐怖袭击伤亡人数备受国际社会高度关注，因此对未来反恐态势的分析主要还是考虑伤亡人数。首先从地区分析伤亡人数，找出伤亡人数比较严重的地区，再从伤亡人数比较严重的地区分析该地区中国家内的伤亡情况，再找出伤亡情况比较严重的国家，分析该国家内不同月份的人员伤亡情况。

4.3.2 时空特性分析

（1）不同地区伤亡人数分析

地区分为 12 类：中美洲和加勒比海地区、南美、东亚、东南亚、南亚、中亚、西欧、东欧、中东和北非、撒哈拉以南的非洲、澳大利亚和大洋洲，对这

12 类地区分别按照 2015 年、2016 年、2017 年计算筛选出总的伤亡人数（见表 4-18），并生成相应的直方图，直方图如图 3.1 所示。

表 4-18 2015-2017 年中东、北非地区国家区域内恐怖袭击伤亡人数

地区	年份		
	2015 年	2016 年	2017 年
北美	133	246	1087
中美洲和加勒比海地区	0	2	5
南美	270	266	244
东亚	157	76	93
东南亚	2023	1826	1629
南亚	17703	16636	16207
中亚	25	46	13
西欧	697	1138	592
东欧	2190	219	284
中东和北非	36430	39507	18032
撒哈拉以南的非洲	13203	9262	9994
澳大利亚和大洋洲	2	1	28

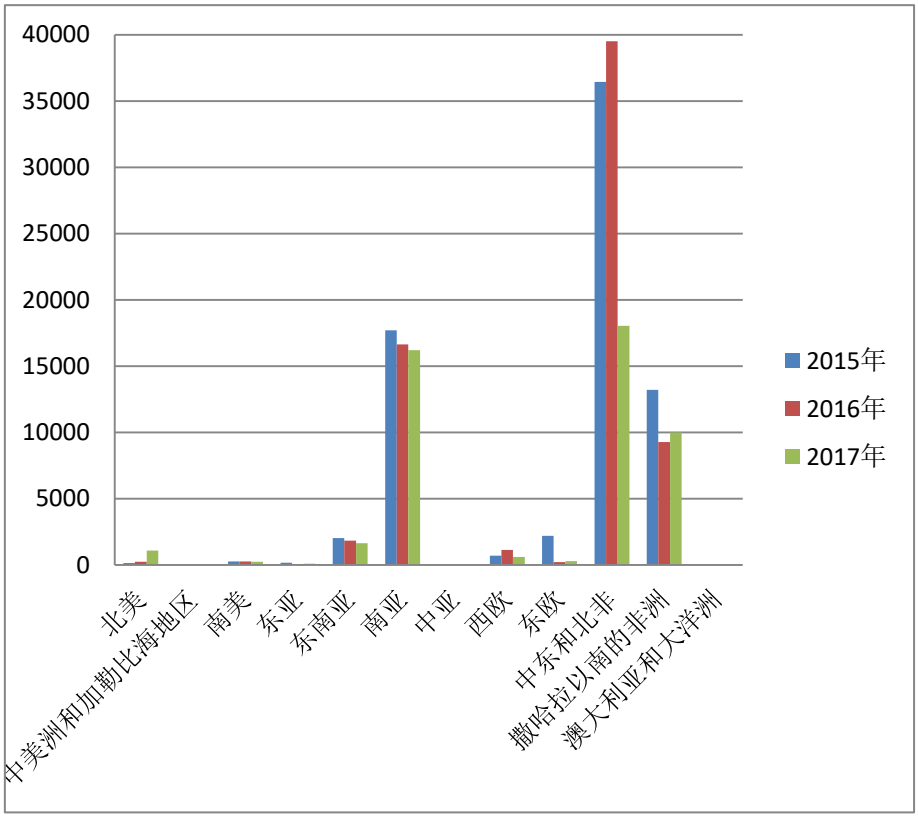


图 4.7 2015-2017 年全球各地区恐怖袭击伤亡人数分布

从直方图 3.1 我们可以直观的看出中东和北非、南亚以及撒哈拉以南的非洲这三个地区 2015-2017 三年间的伤亡人数比较严重，每年伤亡的人数都达到了 10000 人以上，伤亡人数最严重的地区是中东和北非，该地区在 2015 年和 2016

年伤亡人数都超过 35000 人，2016 年伤亡人数接近 40000 人，这一伤亡人数很可怕的，该地区未来的反恐形势要严峻很多，是恐怖袭击的重灾区。

（2）不同国家伤亡人数分析

针对中东和北非、南亚以及撒哈拉以南的非洲地区，分析每个地区内的国家在 2015 年-2017 年伤亡情况。表 4-19 和图 4.8 为中东和北非地区内各国家的伤亡人数情况，我们从图 3.2 中可以清晰看出，Iraq、Syria、Turkey 和 Yemen 这四个国家三年内的伤亡人数比较严重，国家 Iraq 伤亡情况最为严重，三年伤亡人数都达到 10000 人以上，另外三个国家的伤亡人数相对少一点，都低于 5000 人。

表 4-19 2015-2017 年中东和北非地区内各国家的恐怖袭击伤亡人数

Country	年份		
	2015 年	2016 年	2017 年
Algeria	30	11	31
Bahrain	32	3	27
Egypt	1871	1232	1499
Iran	49	20	106
Iraq	20308	25209	10999
Israel	138	191	18
Jordan	18	100	5
Lebanon	458	99	66
Libya	1475	1324	316
Saudi Arabia	434	315	100
Syria	4468	4379	2731
Tunisia	224	97	10
Turkey	1568	3770	548
West Bank and Gaza Strip	356	214	125
Yemen	4750	2541	1446

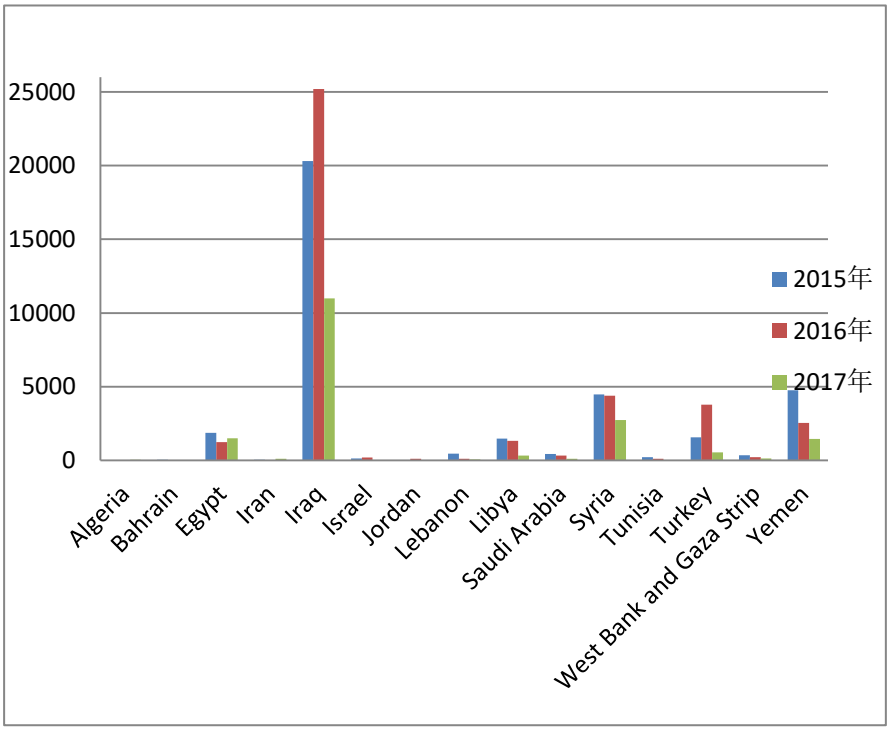


图 4.8 2015-2017 年中东、北非地区内各国家的恐怖袭击伤亡人数

表 4-20 和图 4.9 所示为 2015-2017 年南亚地区内各国家的恐怖袭击伤亡人数，从图 4.9 可以看出国家 Afghanistan 和国家 Pakistan 的伤亡人数比较严重，国家 Afghanistan 的伤亡人数最为严重，虽然这三年的伤亡人数呈下降趋势，但伤亡人数也都在在 12000 人左右。

表 4-20 2015-2017 年南亚地区内各国家的恐怖袭击伤亡人数

Country	年份		
	2015 年	2016 年	2017 年
Afghanistan	12460	12208	11698
Bangladesh	806	165	97
India	1029	1236	1157
Maldives	3	3	1
Nepal	11	20	97
Pakistan	3377	3002	3153
Sri Lanka	17	2	4

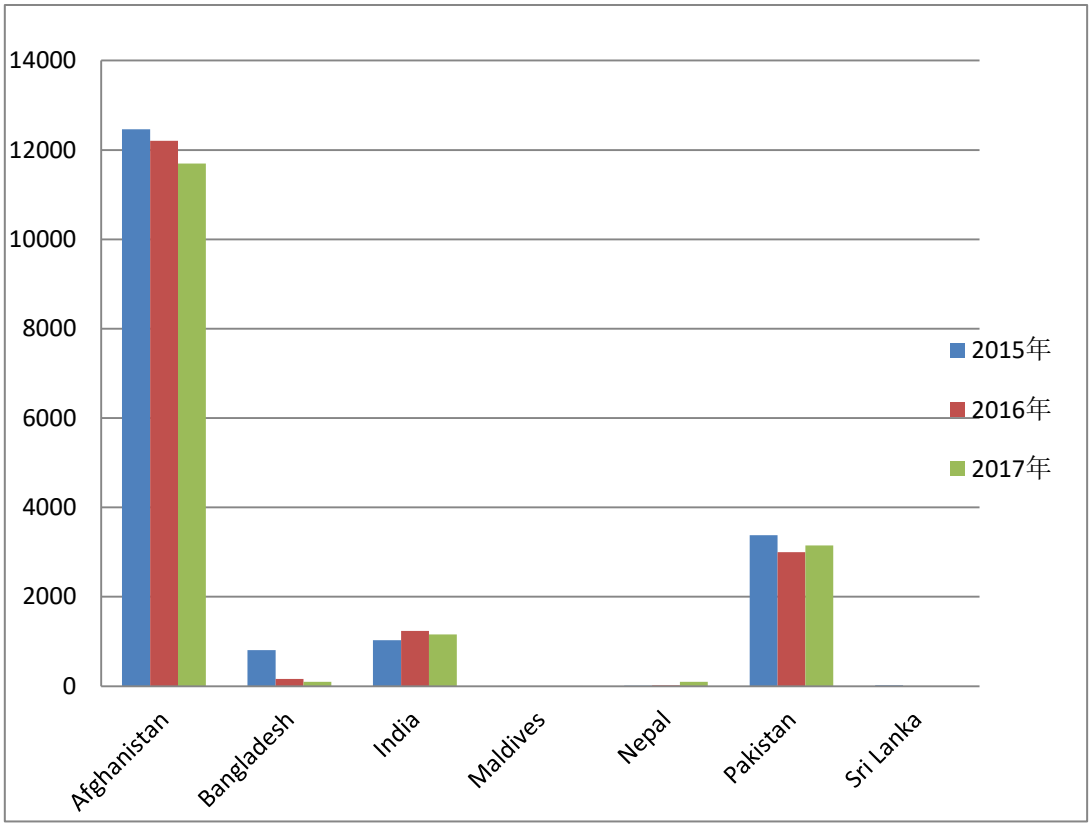


图 4.9 2015-2017 年南亚地区内各国家的恐怖袭击伤亡人数

表 4-21 和图 4.10 所示为 2015-2017 年撒哈拉以南非洲地区内各国家的恐怖袭击伤亡人数，该地区包含的国家比较多，从图 4.10 可以清晰地看出，伤亡人数主要集中在 Nigeria 和 Somalia，在 2015 年 Nigeria 国内出现过一次比较严重的恐怖袭击，发生了 5500 人以上的伤亡，另外的两年的恐怖袭击伤亡人数明显减少了不少，呈下降趋势，Somalia 虽然没发生比较大的人员伤亡的恐怖袭击，但是三年的伤亡人数呈上升趋势，这个国家的恐怖袭击形势比较严峻。

表 4-21 2015-2017 年撒哈拉以南的非洲地区内各国家的恐怖袭击伤亡人数

Country	年份		
	2015 年	2016 年	2017 年
Angola	0	0	416
Burkina Faso	146	85	98
Burundi	434	285	147
Cameroon	1075	567	448
Central African Republic	196	348	569
Chad	673	90	2
Democratic Republic of the Congo	541	531	636
Djibouti	0	0	0
Ethiopia	90	362	80
Gabon	0	0	2
Ivory Coast	24	55	5
Kenya	454	125	204
Liberia	0	0	0

Malawi	0	0	0
Mali	465	390	657
Mozambique	15	145	33
Niger	716	315	210
Nigeria	5758	2403	2300
Rwanda	0	0	10
Sierra Leone	0	1	0
Somalia	1786	2436	3038
South Africa	37	32	28
South Sudan	256	533	836
Sudan	501	173	255
Tanzania	23	5	10
Uganda	5	94	9
Zambia	0	0	0
Zimbabwe	0	0	1

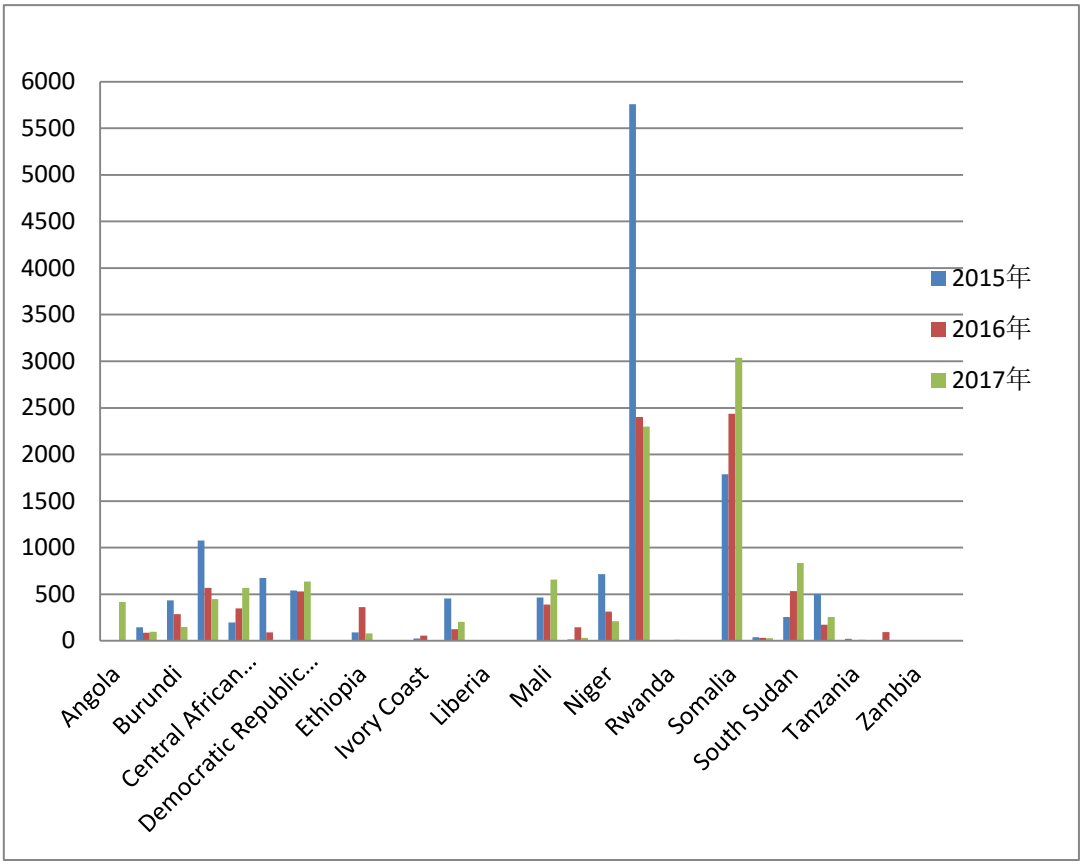


图 4.10 2015-2017 年撒哈拉以南非洲地区内各国家的恐怖袭击伤亡人数

（3）不同月份伤亡人数分析

主要研究中东、北非地区伤亡比较严重的四个国家：Iraq、Syria、Turkey 和 Yemen，在 2015-2017 年三年期间不同月份的人员伤亡情况。如表 4-22 和图 4.11 所示 2015-2017 年 Iraq 国内遭受恐怖袭击的伤亡人数，2015 年每月都发生了伤亡 1500 人左右的恐怖袭击事件，六七八三月份造成的伤亡比较严重；2016 年前半年的人员伤亡人数比较多，后半年相对减少了；2016 年整体上看，恐怖袭击

人员伤亡数都很小，而且呈下降趋势，说明该地区未来的反恐形势比较乐观。

表 4-22　2015-2017 年 Iraq 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	2041	1789	1299
二月	1752	2358	1201
三月	1567	3400	1452
四月	1733	1966	785
五月	1431	2487	1110
六月	1643	2790	1407
七月	2497	2053	835
八月	1797	1359	734
九月	1305	1465	828
十月	1626	2530	459
十一月	1477	1696	506
十二月	1439	1316	383

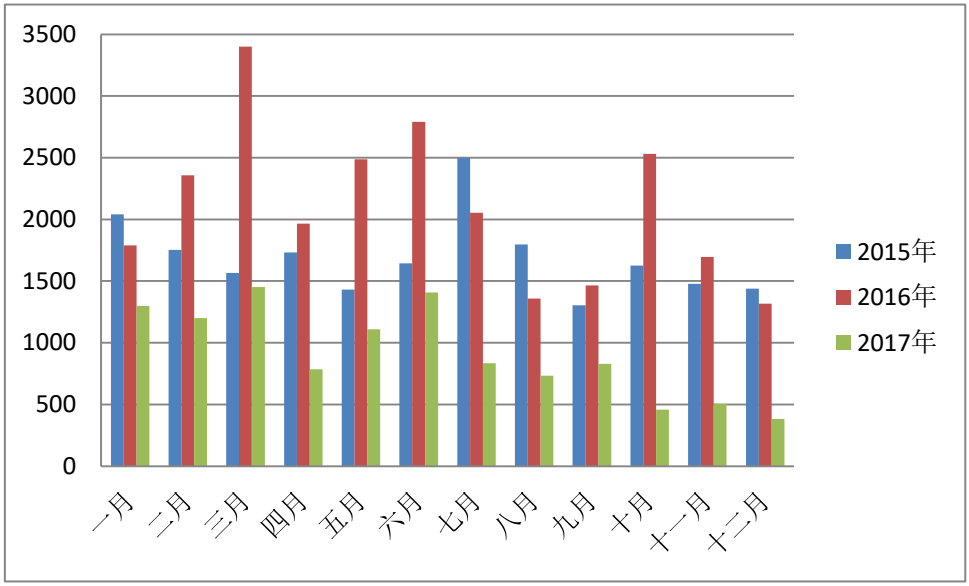


图 4.11 2015-2017 年 Iraq 国内遭受恐怖袭击伤亡人数统计

表 4-23 和图 4.12 所示 2015-2017 年 Syria 国内遭受恐怖袭击的伤亡人数，2015 年伤亡人数每三个月呈上升趋势，2016 年伤亡人数有所减少，2017 年每月的伤亡人数呈波动状态，但整体上比前两年的人员伤亡数要好很多。

表 4-23　2015-2017 年 Syria 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	193	414	317
二月	251	474	403
三月	296	91	163
四月	102	291	311

五月	420	405	331
六月	975	133	118
七月	425	817	189
八月	340	244	68
九月	585	264	105
十月	86	497	366
十一月	260	500	319
十二月	535	249	41

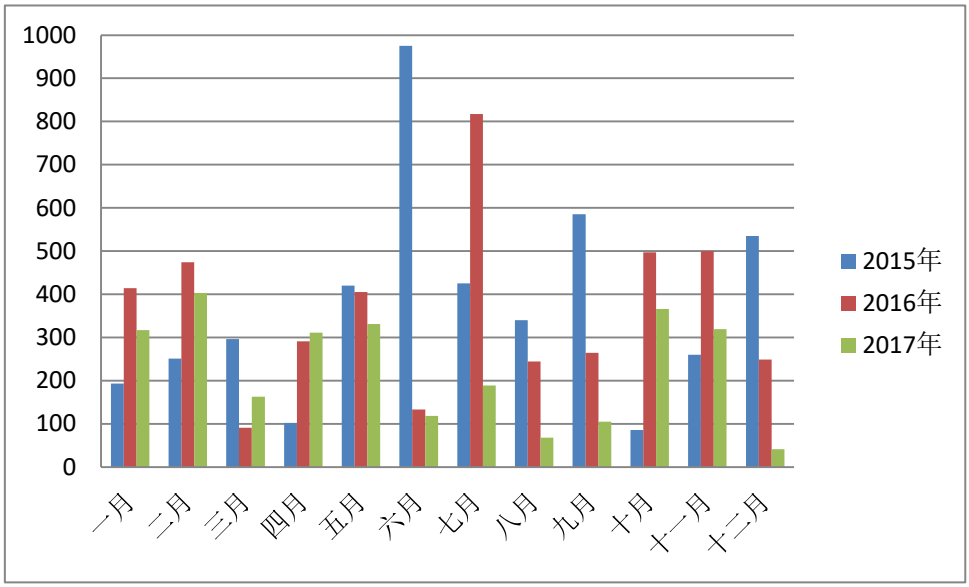


图 4.12　2015-2017 年 Syria 国内遭受恐怖袭击伤亡人数

表 4-24 和图 4.13 所示 2015-2017 年 Turkey 国内遭受恐怖袭击的伤亡人数，2015 年该国家的恐怖袭击比较少，人员伤亡也很少，2016 年每月的伤亡人数呈大幅上升趋势，说明当年该地区的恐怖袭击形势比较严峻，在 2017 年该地区的恐怖袭击人数出现明显下降的趋势，但对于未来的反恐该地区也不能松懈，一出现就会发生很大的人员伤亡恐怖事件，未来更要加强反恐的力度。

表 4-24　2015-2017 年 Turkey 国内遭受恐怖袭击的伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	5	184	134
二月	5	151	19
三月	7	430	17
四月	18	251	50
五月	19	334	28
六月	117	493	69
七月	219	276	65
八月	288	796	40
九月	246	151	44
十月	462	198	46
十一月	93	194	21
十二月	89	312	15

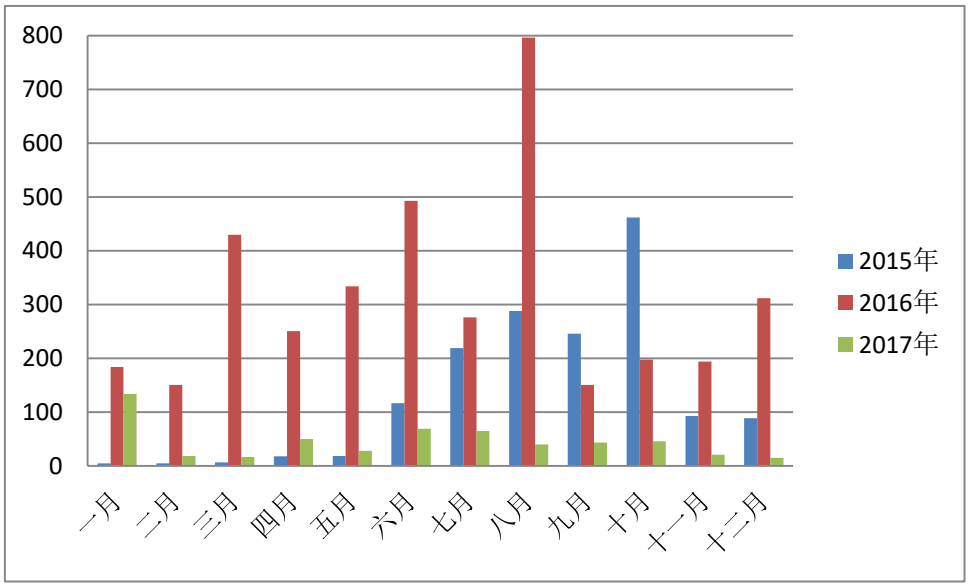


图 4.13　2015-2017 年 Turkey 国内遭受恐怖袭击的伤亡人数

表 4-25 和图 4.14 所示为 2015-2017 年 Yemen 国内遭受恐怖袭击的伤亡人数，从图 4.14 很明显看出 2015 年较于后两年的人员伤亡数比较大，月份主要集中在 3-9 月，后两年的人员伤亡人数明显减少，说明该地区的反恐形势比较乐观。

表 4-25　2015-2017 年 Yemen 国内遭受恐怖袭击的伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	243	264	123
二月	149	159	144
三月	814	146	222
四月	315	243	95
五月	609	332	118
六月	423	249	67
七月	764	277	91
八月	250	209	127
九月	612	115	63
十月	218	101	50
十一月	207	141	259
十二月	146	305	87

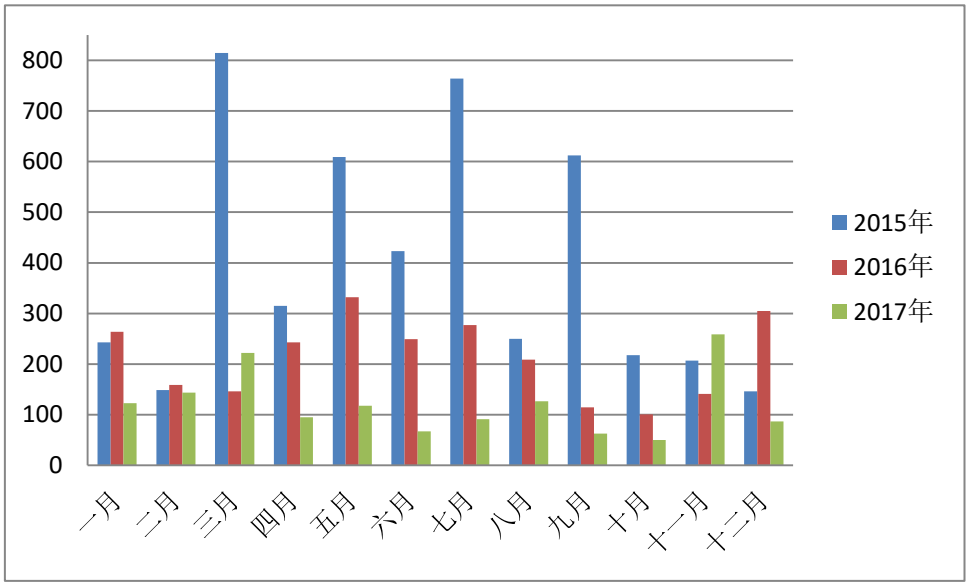


图 4.14　2015-2017 年 Yemen 国内遭受恐怖袭击的伤亡人数

对于南亚地区主要分析了两个伤亡人数比较严重的国家：Afghanistan 和 Pakistan。表 4-26 和图 4.15 所示为 2015-2017 年 Afghanistan 国内遭受恐怖袭击的伤亡人数，从图中可以看出该国家三年中每月的人员伤亡人数都比较大，表明该国家恐怖袭击形势比较严峻，每月都要加强反恐的力度。

表 4-26 2015-2017 年 Afghanistan 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	574	556	691
二月	375	701	633
三月	661	724	845
四月	1244	1503	763
五月	1391	1115	1731
六月	1410	1244	926
七月	1059	1245	1276
八月	1494	1111	1254
九月	2026	1276	795
十月	1115	1531	1240
十一月	527	637	702
十二月	584	565	842

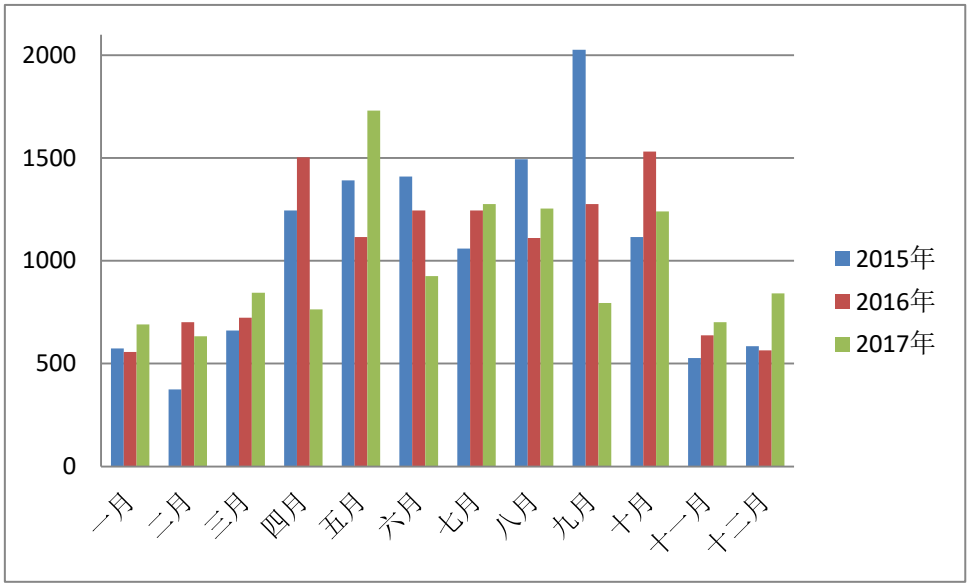


图 4.15　2015-2017 年 Afghanistan 国内遭受恐怖袭击伤亡人数

表 4-27 和图 4.16 所示为 2015-2017 年 Pakistan 国内遭受恐怖袭击的伤亡人数，从伤亡数量上看，该国家每年一月到六月是恐怖袭击伤亡的高发期，其他月份相较于好点，未来该地区前半年要加大反恐力度，后半年也不能放松。

表 4-27 2015-2017 年 Pakistan 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	344	286	207
二月	328	168	722
三月	312	615	289
四月	326	128	134
五月	384	88	172
六月	201	74	461
七月	193	130	224
八月	253	354	207
九月	296	258	137
十月	242	371	210
十一月	311	299	173
十二月	277	77	217

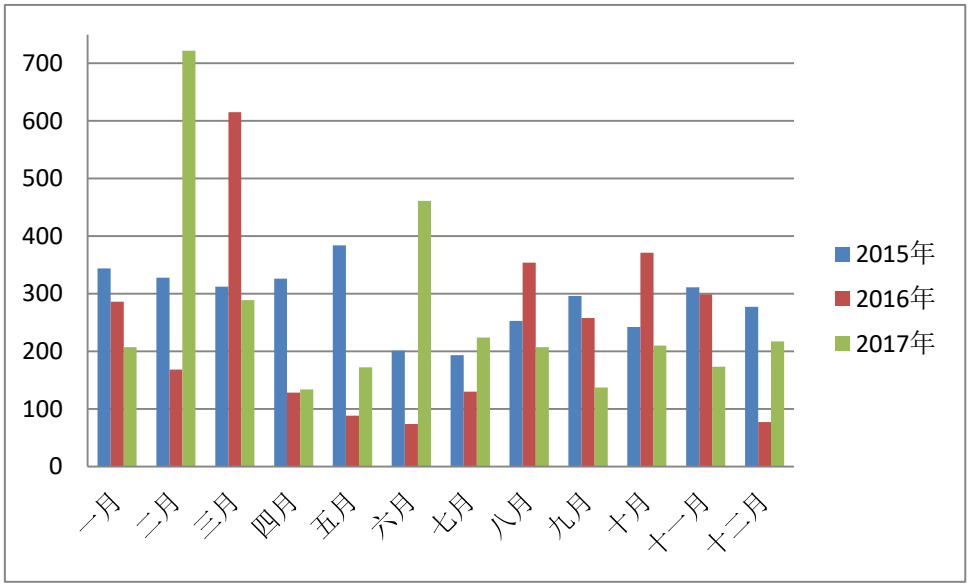


图 4.16 2015-2017 年 Pakistan 国内遭受恐怖袭击伤亡人数

对于撒哈拉以南的非洲地区主要分析了两个伤亡人数比较严重的国家：Nigeria 和 Somalia。表 4-28 和图 4.17 所示为 2015-2017 年 Nigeria 国内遭受恐怖袭击的伤亡人数，该地区 2015 年每月的恐怖袭击伤亡人数明显比较多，其他两年每月的伤亡情况比较缓和，每年的第一季度和第四季度伤亡人数较多，未来该地区要加强第一季度和第四季度的反恐力度。

表 4-28 2015-2017 年 Nigeria 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	423	386	213
二月	666	276	96
三月	399	247	133
四月	294	178	181
五月	355	134	199
六月	617	76	180
七月	815	129	203
八月	280	67	205
九月	512	132	210
十月	577	139	137
十一月	341	284	348
十二月	479	355	195

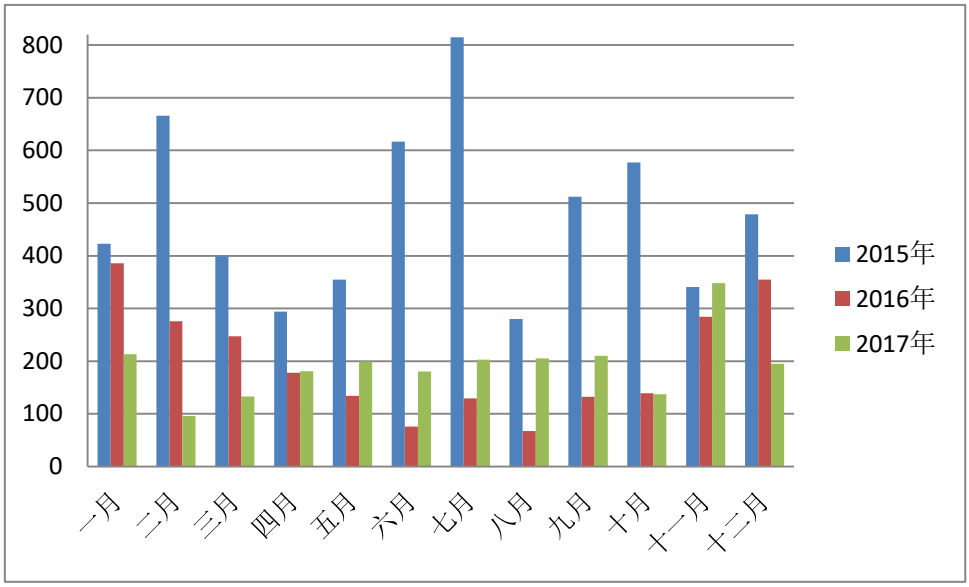


图 4.17 2015-2017 年 Nigeria 国内遭受恐怖袭击伤亡人数

表 4-29 和图 4.18 所示为 2015-2017 年 Somalia 国内遭受恐怖袭击的伤亡人数，从柱状图中看出，2017 年该国家伤亡人数出现上升，尤其在 10 月份，未来该地区在一月份、六月份和十月份要加强恐怖打击力度，防范恐怖袭击。

表 4-29 2015-2017 年 Somalia 国内遭受恐怖袭击伤亡人数

月份	年份		
	2015 年	2016 年	2017 年
一月	78	150	416
二月	207	257	194
三月	136	293	100
四月	109	150	200
五月	48	139	123
六月	193	171	295
七月	273	185	168
八月	265	326	110
九月	148	93	191
十月	33	193	1104
十一月	166	231	55
十二月	130	248	82

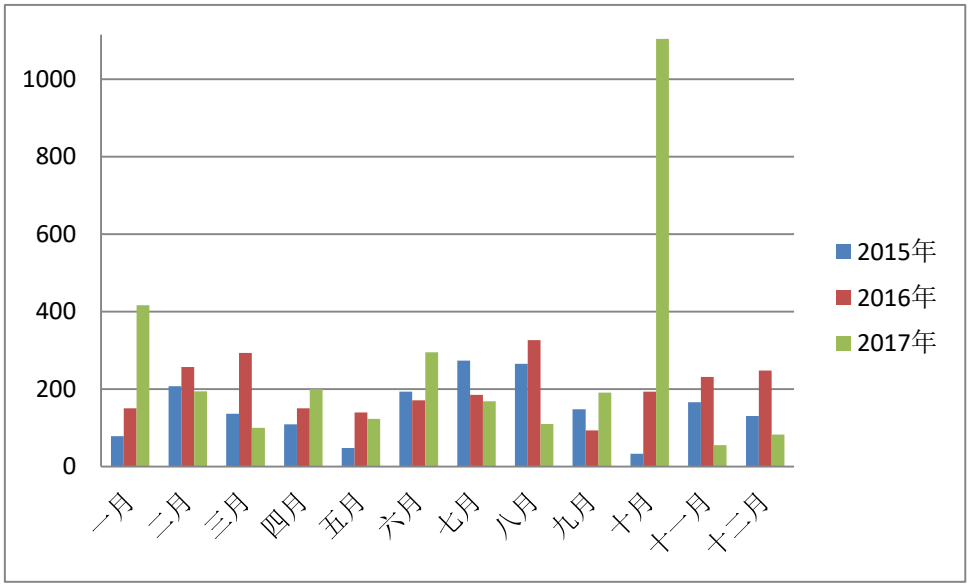


图 4.18 2015-2017 年 Somalia 国内遭受恐怖袭击伤亡人数

4.3.3 蔓延特性分析

分析 2015-2016 年近三年全球所遭受的恐怖袭击事件，按照年份分类筛选出相对应的经纬度坐标值，使用数据处理软件 R 软件根据经纬度坐标值绘制出恐怖事件全球蔓延的地理图，如图 4.19 所示，黄色的表示为 2015 年的，绿色表示为 2016 年的，蓝色表示为 2017 年的，从图上我们可以更加直观的看出恐怖事件的发生还主要集中在中东和北非、南亚以及南非地区，也说明了这三个地区的恐怖袭击活动比较严重，蓝色区域开始包含东南亚、西欧、东欧等地区，中东和北非的恐怖活动有往南非、北美、东欧、东南亚区域蔓延的趋势。

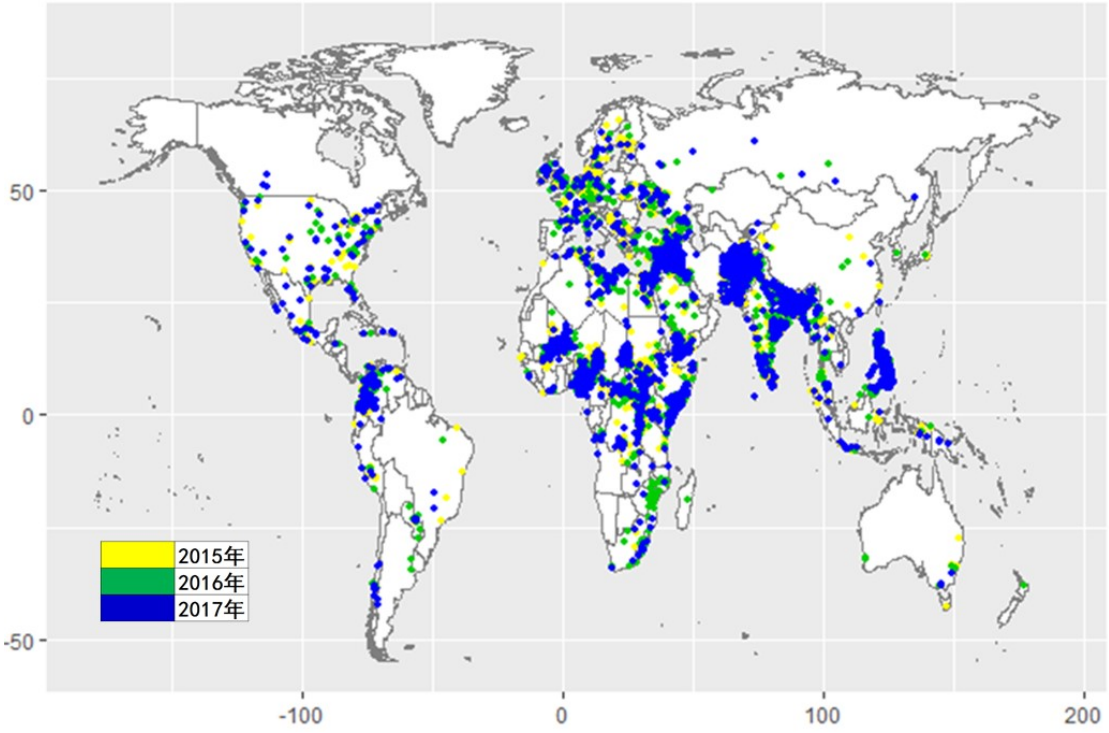


图 4.19 2015-2017 年全球发生恐怖事件的分布图

4.3.4 级别分布分析

结合问题一的等级划分求解，筛选出 2015-2017 年的经纬度坐标值，运用 R 软件将问题一划分好的 5 类等级按照经纬度坐标值画出全球分布矢量图，如图 4.20 所示，红色代表一级等级，蓝色代表二级等级，青色代表三级等级，绿色代表四级等级，黄色代表五级等级，从图中可以清楚看出代表一级等级的红色主要集中在中东、中亚等地区，这些地区恐怖分子活动比较猖獗，所制造的恐怖事件都是比较严重的，该地区的反恐形势还是极其严峻的。

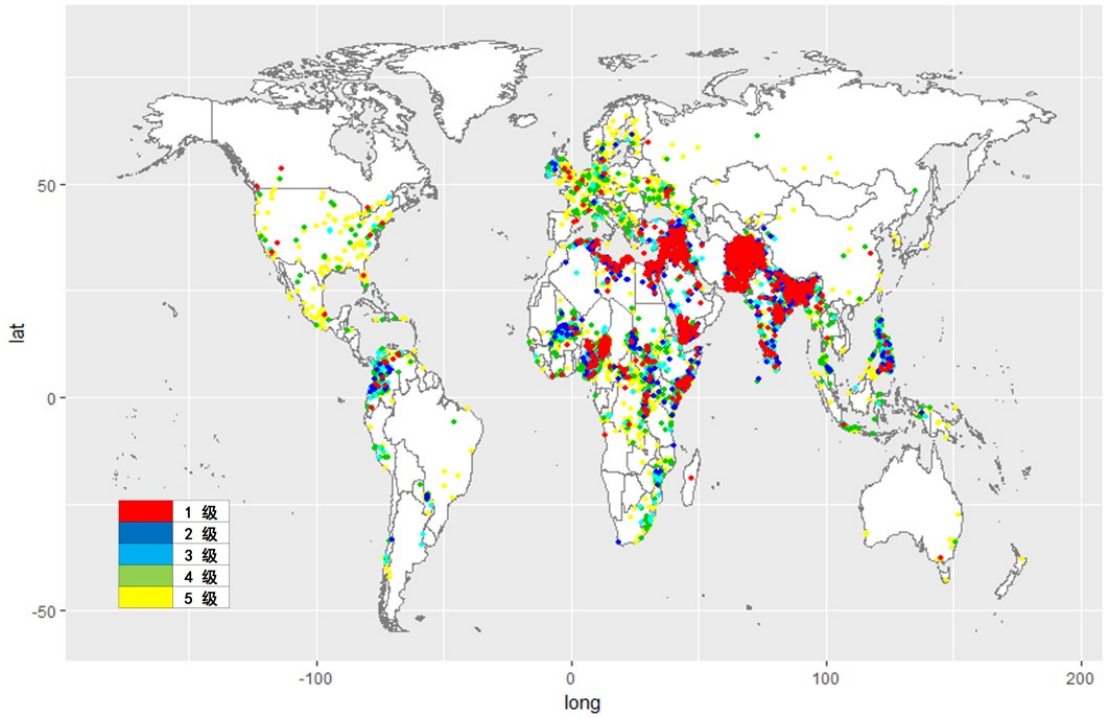


图 4.20 2015-2017 年全球恐怖事件等级分布图

4.3.5 主要原因分析

恐怖袭击发生的主要原因分析：

1、近代大国的政治局势稳定，导致各国对资源的需求产生竞争。从而更多的以各种形态介入资源地区制造纷争。导致某些地区的长期不发达，战乱，让宗教或极端势力能在当地人民产生被侵略的感想中发展。（例如伊拉克、阿富汗）

2、由于军队的入侵或者进驻，导致一系列的所谓民主输出，西方精神输出，与当地的习俗、宗教产生抵触。（泛伊斯兰地区）。

3、由于进入近代，各国、各民族都大力提倡民族思想教育，导致全世界的人类对民族的认同感，认知能力大幅度提高。产生各样的民族独立或民主行动但是由于现世界格局稳定，各国可以抽手出来控制，镇压，由于军事能力上的差距，导致出现各样的武装袭击。（例如北爱尔兰、西班牙）。

4、历史遗留的民族问题。例如，（车臣）前苏联没能在立国的最初改善车臣人的政治地位与融合，直接就是以暴力的方式解决，后来政策调整，又拿国家的钱去养车臣但车臣觉得是国家的补偿，并不感谢，导致国内的主流民族觉得自己在倒贴却没能有好结果，从而加速两民族对立。由于民族上的强弱悬殊，在正面战场上车臣人的完败结局，最后车臣人只能转到三的这种办法去抗争。

5、由于现代的通信传媒发达，从而使得恐怖份子的任何行为，会被马上放大，产生一

个模范效应。一方面，让恐怖组织更出名，资金的来源，人员的募集更方便。另外在一些政治斗争上更容易获得好处或者打击现政府。从而出现恶性循环，组织越小，越搞袭击，越袭击，利益越大、好处越多，成名越快。成功的袭击，其手法，工具，思路被完全披露，邮让其他的组织更容易学习，或受到启发，引发了更多的恐怖袭击事件。

4.3.6 对未来反恐形势的分析及应对措施

综合上述的分析可以看出，恐怖分子的活动主要还是积聚于中东与北非、南亚、撒哈拉以南的非洲，这些地区主要是以“伊斯兰国”为代表的国际恐怖主义势力，近几年国际社会一直在联合打击“伊斯兰国”恐怖主义势力，使其力量得到一定的削弱，在这种压力下，他们为了求生存，积蓄力量，开始往外蔓延其他地区，如南非、北美、东欧、东南亚，进而分散被打击的力量，全球的恐怖形势还是很严峻的，为此本文针对应对未来反恐态势所采取的措施提出以下几点建议：

（1）集中力量重点打击恐怖主义的重灾区。恐怖主义的重灾区是恐怖主义极端组织思想传播和恐怖人员输送的主要发源地，通过打击重点区域内的重点恐怖分子聚集的城市，打掉他们的根据地，进而有助于削弱中心城市对其周边城市的恐怖渗透的能力，中东与北非地区重点打击的城市包括：Iraq、Syria、Turkey 和 Yemen，南亚地区重点打击城市有：Afghanistan 和 Pakistan，撒哈拉以南的非洲地区重点打击的城市有：Nigeria 和 Somalia。

（2）加强国际间的协同合作。以“伊斯兰国”代表的极端恐怖组织开始慢慢的渗透到其周边比较安定的地区，发展与壮大自己的恐怖组织力量，因此国际间要加强紧密合作，对于恐怖组织的相关信息要及时共享，阻断恐怖组织渗透的发展，共同歼灭有预谋的、跨国家区域的恐怖组织活动，反对将恐怖主义与特定国家、民族和宗教挂钩，推动不同文明对话交流，加强人文沟通，增进政治互信，营造有利于开展反恐国际合作的大环境。

（3）提高民众对于恐怖主义的认识以及自己的安全防范意识。普通民众需要对恐怖主义有一定的认知，恐怖主义有些什么特点、危害性等一些常识，可以去做一些自己力所能及的事情，提高自己的安全意识，危机意识，民众要保持一定的警觉性，一旦发现恐怖主义苗头或者蛛丝马迹，就向安全部门通报，及时制止恐怖事件的发生。

（4）各国家要加强对互联网的监管与审核力度，防范恐怖主义思想在信息化时代的快速传播。恐怖组织滥用当今世界全球化、信息化的飞速发展，不断传播暴力极端思想和制暴技术，荼毒全球网民的思想，变本加厉危害人类社会，必须要对其进行相应的遏制。

4.4 问题四：数据的进一步利用

4.4.1 题目分析

目前在电子商务领域，实体商务领域，电信网络等比较依赖用户基数的商业领域，为了解决潜在的客户流失问题，常采用基于统计学的方法如贝叶斯分类器、聚类分析、决策树、回归等，建立客户流失状态模型[14]，模型可精准地为客户提供有效的产品服务系统整体解决方案，达到维持客户吸引力的目的[15]。

其中，一种改进粒子群算法与支持向量机相结合的客户流失预测方法(IPSO-SVM)[16]，在客户流失模型中具有优秀的有效性和可行性，本文利用此方法，建立对已经确定的恐怖组织，选取评价指标，如袭击月份（month）、袭击

地区（region）和国家（country）、袭击目标类型（target type）、袭击类型（attack type）、使用武器类型（weapon type）、动机（motive）等信息，建立恐怖组织“客户流失预测模型”，以附件1恐怖袭击历史记录信息为基础，对潜在的恐怖组织状态进行判断。根据预测的结果，特定地区政府对以上评价指标定性或定量的采取相应的措施，提高恐怖组织的“撤离率”，以达到降低该地区遭受恐怖袭击几率的目的。

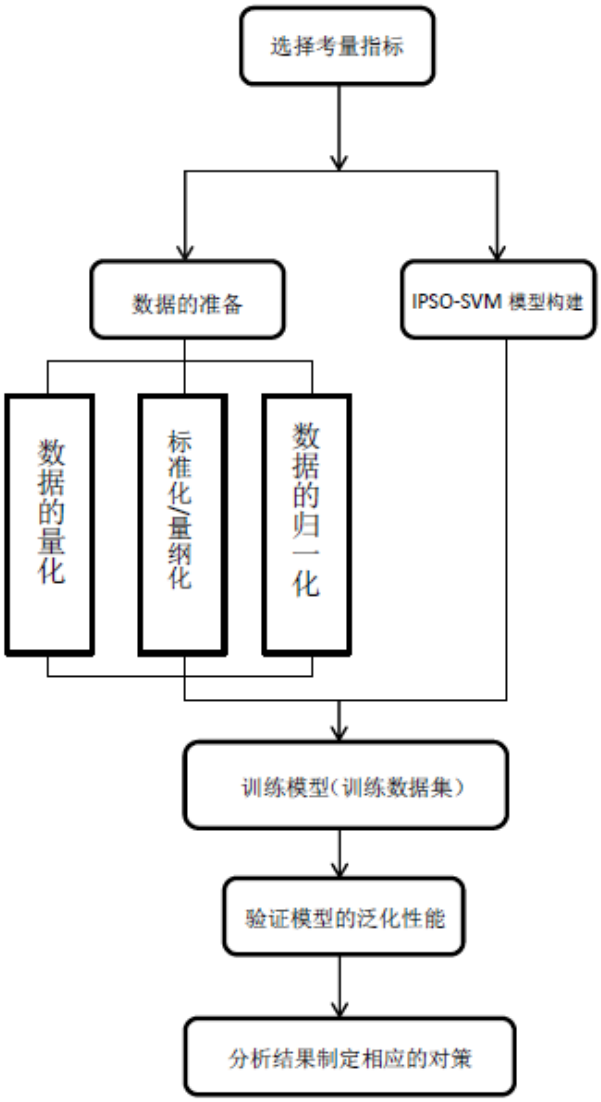


图4.21 技术路线总图

技术路线如上图所示。主要过程如下：首先从恐怖组织作案历史记录中抽取主要评价指标，然后通过数据挖掘方法挖掘出具有潜在“撤离倾向”的组织，最后政府组织根据挖掘出结果，及时制定出具体的影响策略和当地武装力量提升策略，增加恐怖组织的“撤离率”。其中，恐怖组织“撤离”预测模型流程图如下：

图4.22 “撤离”预测模型流程图

4.4.2 数据准备与处理

从某个地区在某一时间段内发生恐怖袭击事件的状况数据进行整理与分析，从客户历史数据中抽取7个恐怖袭击的评价指标：袭击月份（month）、袭击地区（region）和国家（country）、袭击目标类型（target type）、袭击类型（attack type）、使用武器类型（weapon type）、动机（motive）。初始化数据集中有114185个样本，筛选出已经确定恐怖的事件，选取2015-2017年的数据作为训练数据集。然后按照前文提到的方法对各个指标的数据进行量化（见下表），其中流失状态[1]表示某恐怖组织袭击了该地区后未撤离或隐匿，[2]表示袭击后撤离或者隐匿。预测“撤离”计算指标矩阵如下表所示。

表（1）初始数据集（部分）

事件编号	年份	袭击月份	地区国家	袭击目标类型	攻击类型	武器类型	犯罪集团的名称	撤离状况
201412220095	2015	0.08	0.0050	0.0230	0.3880	2.9246	uslim extremist	2
201501010003	2015	0.08	0.2440	0.0190	0.2730	4.7877	of Benghazi Re	1
201501010008	2015	0.08	0.2440	0.1280	0.2630	4.7877	of Iraq and the	2
201501010009	2015	0.08	0.2440	0.2840	0.3880	3.3869	of Iraq and the	2
201501010014	2015	0.08	0.0730	0.1680	0.3880	3.3869	t (OPM-Organis	1
201501010025	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010026	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	1
201501010027	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010028	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010030	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010031	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010032	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010033	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010034	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	2
201501010035	2015	0.08	0.0050	0.1630	0.2730	4.7877	sk People's Rep	1
201501010036	2015	0.08	0.0050	0.1630	0.3880	4.7877	sk People's Rep	1
201501010037	2015	0.08	0.0050	0.1630	0.3880	3.3869	sk People's Rep	1
201501010038	2015	0.08	0.0830	0.0230	0.2730	4.7877	Boko Haram	2
201501010039	2015	0.08	0.0830	0.0370	0.3880	2.9246	Boko Haram	2
201501010067	2015	0.08	0.0050	0.2840	0.2730	4.7877	sk People's Rep	2
201501010069	2015	0.08	0.2550	0.2840	0.3880	3.3869	of India - Mao	2
201501010071	2015	0.08	0.2440	0.1680	0.2730	4.7877	he Arabian Pen	2
201501010076	2015	0.08	0.0830	0.1630	0.3880	3.3869	Al-Shabaab	2

图4.23 初始数据集部分截图

表4-30预测“撤离”计算指标矩阵

恐怖组织状态	预测“撤离”组织的数目	预测“撤离”组织的数目
实际“撤离”	A	B
实际“非撤离”	C	D

模型评价指标为模型准确率、命中率、覆盖率与提升系数，评价标准由预测流失计算指标矩阵获得，并有如下定义：

$$\text{准确率} = \frac{A + D}{A + B + C + D}$$

$$\text{命中率} = \frac{A}{A + B}$$

$$\text{覆盖率} = \frac{A}{A + B}$$

$$\text{提升度} = \frac{\text{命中率}}{\text{测试数据集中的客户流失率}}$$

4.4.3 模型的建立与求解

首先，构造训练集为：{(x,y)…}，x是输入样本数据，y是样本类别。SVM 可将样本通过非线性函数*f(x)*映射射到高维度空间G，并在高维空间中进行线性回归，其线性回归函数为：

$$g(x) = \omega \times f(x) + b \tag{4-15}$$

式中，权重向量ω，偏置为b。

根据结构风险最小化原则，可将支持向量机进行估计回归的问题转化为如下优化问题：

$$\begin{cases} \min J = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\vartheta_i^* + \vartheta_i) \\ s.t. \quad y_i - \omega * f(x) - b \le \varepsilon + \vartheta_i \\ \quad\quad \omega * f(x) + b - y_i \le \varepsilon + \vartheta_i^* \\ \quad\quad \vartheta_i^* \ge 0, \vartheta_i \ge 0, i = 1,2,3\ldots,n \end{cases} \tag{4-16}$$

式中，$\frac{1}{2}\|\omega\|^2$表示的是模型的复杂度，其值越大，模型越复杂；ϑ_i与ϑ_i^*为松弛因子；惩罚参数为C，用来衡量模型的复杂度与经验风险。

为了进一步简化上述模型，引入拉格朗日函数，将其转化为二次优化问题，则有：

$$\begin{aligned} &L(\omega, b, \vartheta_i, \vartheta_i^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) \\ &= \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\vartheta_i^* + \vartheta_i) - \sum_{i=1}^{n}\alpha_i * \left(\vartheta_i + \varepsilon - y_i + g(x_i)\right) \\ &- \sum_{i=1}^{n}(\vartheta_i * \gamma - \vartheta_i * \gamma_i^*) \end{aligned} \tag{4-17}$$

式中，α_i, α_i^*均为拉格朗日乘子。

则得到的支持向量机预测模型为：

$$g(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) * f(x_i, x) + b \tag{4-18}$$

在进行非线性预测时，最关键的是选择合适的核函数以及确定相对应的参数。常用的核函数有Sigmoid核函数、多项式核函数、径向基核函数等。核函数及其参数通常情况下是通过不断的进行实验，然后经过对比分析的方法获取。但是在大多数情况下，选取径向基核函数进行预测的性能较好[17]，所以本文采用径向基核函数。径向基核函数可表示为：

$$K(x_i, x_j) = \exp\left(-\delta * \|x_i - x_j\|^2\right), \delta > 0 \tag{4-19}$$

恐怖组织“流失”预测模型中需要对SVM参数进行优化，目的是提高预测的准确度。本文采用将预测准确度(T)作为优化目标函数，因此，PSO对SVM 参数优化目标函数为：

$$\begin{aligned} &\mathrm{MaxT}(\mathrm{C},\delta) \\ &\text{s.t.} \\ &\begin{cases} C_{min} \le C \le C_{max} \\ \delta_{min} \le \delta \le \delta_{max} \end{cases} \end{aligned} \tag{4-20}$$

模型中参数(C,δ) 的选取决定了是否能够精确预测。在通常情况下，将整理好的数据分为两组，其中一组用于训练以保证预测精度，而交叉验证的方法是该方法的改进方法。通过交叉验证的方法搜索到最优的参数(C,δ)，选取精度最高的一组参数作为最优的参数。

则最终的支持向量机的预测模型为:

$$\mathrm{g(x)} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) * \exp\left(-\frac{\|x_i - x_j\|}{2*\delta^2}\right) + b \tag{4-21}$$

采用基于IPSO-SVM的恐怖组织“撤离”预测模型求解流程如下:

① 对数据集进行归一化处理，将全部信息映射到［0，1］区间内；

② 初始化各个输入参数，基于*Sobol*序列初始化粒子群，每个粒子代表一组参数(C,δ)；

③ 利用SVM进行学习和训练，并通过交叉验证的思想计算目标函数值。更新粒子的位置和速度；

④ 计算每个粒子目标函数值，并更新全局最优解*gbest*和局部最优解*pbest*；

⑤ 如果达到迭代次数，保存d全局最优粒子位置，否则转至步骤(3)；

⑥ 将最优位置对应的SVM参数构建产品服务系统客户流失预测模型。

4.4.4 模型结论分析与对策建议

根据IPSO-SVM的预测结果分析，可以选取出按重要程度划分的评价指标，根据该地区的恐怖组织“撤离率”大小，当地政府给出相应的对策。例如，某个地区的分析结果中某个月份恐怖组织的“撤离率”较低，政府可以采取对该地区加强军事力量的输入，通过媒体发布消息等政策以威慑当地的恐怖组织；或者政府可以对某个地区受到的攻击类型、武器类型正对性的做出布防和准备，就可以降低该地区的损失程度。

五、模型评价与改进

问题一：本文建立的依据危害性对恐怖袭击事件分级模型，同时使用因子分析法和 CRITIC 赋权值法对题目提供的恐怖袭击事件记录数据进行分析，得出所有事件的综合得分然后按照事件的危害程度进行排名，最后采用 DBCLASD 聚类方法将事件得分按危害程度分为了 1-5 级。模型 1 采用因子分析法可以有效地提取数据的公因子，对数据进行了降维处理，减轻了计算的难度，同时使用更具客观性的 CRITIC 赋权值法进行计算，提高了事件评价指标的准确性。但是，该模型在评价指标选取时不能涵盖所有的事件特点，数据还具有一定的片面性，会影响结果的准确性。故，在选取评价指标方面还有待优化。

问题二：依据事件特征发现恐怖袭击事件制造者的模型，首先把 2015-2016 年已知的作案对象的事件记录分为一个类别，那么未知作案人的事件为另一类，首先对未知恐怖组织或个人的样本与已知恐怖组织或个人样本的相似性进行训

练，使用的是 Pearson 算法计算相关系数，然后根据问题一得到的排名筛选出2016-2016 年危害程度得分最大高的 5 个组织或个人标记为 1-5 号嫌疑人，对题目所给事件进行相关性检验，按相关性大小匹配嫌疑程度。比较遗憾的事，由于时间有限，提高模型的准确性未使用更精确的 RBF 神经网络分类、SVDD 算法等进行验证提高模型的准确度。

问题三：从时空特性、蔓延特性、级别分布和主要原因四个方面来对未来反恐态势进行相关分析，时空特性主要是基于国际社会比较关注的伤亡人数作为主要的研究对象，可能比较单一，因此结合蔓延特性和级别分布的全球分布图，可以更加直观的分析未来反恐形势。

问题四：借鉴了其他商业销售领域常用的客户流失预测模型原理，建立恐怖组织“撤离预测模型”，使用改进粒子群算法与支持向量机相结合的客户流失预测方法（IPSO-SVM）对数据进行训练，分析出各个评价指标的影响程度，知道当地政府制定出针对这些指标变化的相应对策。该模型还未在恐怖袭击事件预测进行实际的验证，还需要更多的数据和更长的时间进行检验。

六、参考文献

[1] 荣钰湘.恐怖主义的认识分歧与解决之路[D].北京：中国政法大学，2016.
[2] 陈桂香.加强地铁安全造福千家万户[J].中国安防，2010（9）：104-107.
[3] 李勇男，梅建明，秦广军.反恐情报分析中的数据预处理研究[J].情报科学，2017（11）：103-107.
[4] 谭啸.我国反恐情报的搜集与利用[D].南京：南京大学，2015.
[5] 位珍珍.后 911 时代恐怖主义的 GTD 数据分析[J].情报杂志，2017,36(7):10-15.
[6] 李国辉.全球恐怖袭击时空演变及风险分析研究[D],中国科学技术大学.2014.05.
[7] A Clauset, FW Wiegel. A Generalized Aggregation-Disintegration Model for the Frequency of Severe Terrorist Attacks[J].Journal of Conflict Resolution, 2010, 54 (1) :179-197.
[8] 刘勇，陆愈实.基于因子分析法的道路安全交通安全评价与决策研究[J].安全与环境工程，2009，16（06）：112-114.
[9] 张立军，张潇.基于改进 CRITIC 法的加权聚类方法[J].统计与决策，2015,22:65-68.
[10] 鲍学英，李海连，王起才.基于灰色关联分析和主成分分析组合权重的确定方法研究[J].数学的实践与认识,2016,46(9):129-134.
[11] 宋浩远.基于模型的聚类方法研究[J].重庆科技学院学报（自然科学版）2008,10(3):71-73.
[12] 杨帆，冯翔，阮羚. 基于皮尔逊相关系数法的水树枝与超低频介损的相关性研究[J].高压电路,2014,50(6):21-25.
[13] 王涓，吴旭鸣，王爱凤. 应用皮尔逊相关系数算法查找异常电能表用户[J].电力需求测管理,2014,16(2):52-54.
[14] 赵宇，李兵，李秀，等. 基于改进支持向量机的客户流失分析研究[J].计算机集成制造系统，2007,13(1) : 20－207.
[15] 王观玉，郭勇.支持向量机在电信客户流失预测中的应用研究[J].计算机仿真 2011，28(4) : 115－118.
[16] 张伟，师奕兵，周龙甫，等.基于改进粒子群算法的小波神经网络分类器［J］.仪器仪表学报，2010,31(10) :2203－2209.
[17] 卓涛. 基于粒子群优化支持向量机的电子商务客户流失预测模型［J］. 农业网络信息，2014(6) : 88－91.

七、附录

附录一、R 语言程序

1. 问题三：绘制 2015-2017 年全球发生恐怖事件分布图的 R 语言代码

```
setwd("E:/data2")#设置工作路径

require(openxlsx)
rm(list=ls())

data1<-read.table("蔓延 2015.csv",header=TRUE,sep=",")  #导入 2015 恐怖事件 csv 文件
visit1.x<-data1$J
visit1.y<-data1$W #数据准备

data2<-read.table("蔓延 2016.csv",header=TRUE,sep=",")  #导入 2016 恐怖事件 csv 文件
visit2.x<-data2$J
visit2.y<-data2$W #数据准备

data3<-read.table("蔓延 2017.csv",header=TRUE,sep=",")  #导入 2017 恐怖事件 csv 文件
visit3.x<-data3$J
visit3.y<-data3$W #数据准备

library(ggplot2)

library(ggmap)

library(sp)

library(maptools)

library(maps)

mp1<-NULL #定义一个空的地图

mapworld<-borders("world",colour = "gray50",fill="white") #绘制基本地图

mp1<-ggplot()+mapworld+ylim(-55,85)#利用 ggplot 呈现，同时地图纵坐标范围从-55 到 85

mp2<-mp1+geom_point(aes(x=visit1.x,y=visit1.y,size=data1$number),color=7)+scale_size(range
=c(1,1))
mp3<-mp2+geom_point(aes(x=visit2.x,y=visit2.y,size=data2$number),color=3)
mp4<-mp3+geom_point(aes(x=visit3.x,y=visit3.y,size=data3$number),color=4)
```

#绘制带点的地图，geom_point 是在地图上绘制点，x 轴为经度信息，y 轴为纬度信息，size 是将点的大小按照收集的个数确定，color 为暗桔色，scale_size 是将点变大一些

mp5<-mp4+theme(legend.position = "none") #将图例去掉

mp5 #地图呈现

2. 问题三：绘制 2015-2017 年全球恐怖事件级别分布图的 R 语言代码

```
setwd("E:/data2")#设置工作路径

require(openxlsx)
rm(list=ls())

data1<-read.table("1.csv",header=TRUE,sep=",")   #导入等级 1csv 文件
visit1.x<-data1$J
visit1.y<-data1$W #数据准备

data2<-read.table("2.csv",header=TRUE,sep=",")   #导入等级 2csv 文件
visit2.x<-data2$J
visit2.y<-data2$W #数据准备

data3<-read.table("3.csv",header=TRUE,sep=",")   #导入等级 3csv 文件
visit3.x<-data3$J
visit3.y<-data3$W #数据准备

data4<-read.table("4.csv",header=TRUE,sep=",")   #导入等级 4csv 文件
visit4.x<-data4$J
visit4.y<-data4$W #数据准备

data5<-read.table("5.csv",header=TRUE,sep=",")   #导入等级 5csv 文件
visit5.x<-data5$J
visit5.y<-data5$W #数据准备

library(ggplot2)

library(ggmap)

library(sp)

library(maptools)

library(maps)

mp<-NULL #定义一个空的地图
```

```
mapworld<-borders("world",colour = "gray50",fill="white") #绘制基本地图

mp<-ggplot()+mapworld+ylim(-55,85)#利用 ggplot 呈现，同时地图纵坐标范围从-55 到 85

mp2<-mp+geom_point(aes(x=visit5.x,y=visit5.y,size=data5$number),color=7)+scale_size(range=
c(1,1))
mp3<-mp2+geom_point(aes(x=visit4.x,y=visit4.y,size=data4$number),color=3)
mp4<-mp3+geom_point(aes(x=visit3.x,y=visit3.y,size=data3$number),color=5)
mp5<-mp4+geom_point(aes(x=visit2.x,y=visit2.y,size=data2$number),color=4)
mp6<-mp5+geom_point(aes(x=visit1.x,y=visit1.y,size=data1$number),color=2)
#绘制带点的地图，geom_point 是在地图上绘制点，x 轴为经度信息，y 轴为纬度信息，size
是将点的大小按照收集的个数确定，color 为暗桔色，scale_size 是将点变大一些

mp7<-mp6+theme(legend.position = "none") #将图例去掉

mp7 #将地图呈现出来
```

附录二、

附件：相关数据文件